AD-A236 763

(2)

N PAGE

Form Approved
OMB No. 0704-0188

AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June, 1989	Final report covering conference:5/89

4. TITLE AND SUBTITLE
"Physiology-Psychophysies Session at the Spring, 1989 ASA Meeting"

5. FUNDING NUMBERS
G-AFOSR-89-0314
PE- 61102F
2313
A6

6. AUTHOR(S)
Christopher W. Turner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Acoustical Society of America
500 Sunnyside Blvd.
Woodbury, New York 11797

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 91 0166

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr Genevieve Haddad
AFOSR/NL
Building 410
Bolling AFB, DC 20332

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Public availability-distributed to conference participants and in response to mail requests.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words) A special session of the Spring 1989 Acoustical Society of America Meeting was sponsored by this grant. The special session concerned the interactions between auditory psychophysics and physiology research. Six invited papers were presented in the morning session and a contributed-paper poster session was held in the afternoon. The topics addressed included 1) meaningful comparisons between psychophysical results for discrimination of sounds and possible physiological counterparts, 2) modeling the responses of the auditory nervous system to account for psychophysical data and 3) new techniques for the collection of physiological data. The invited papers were collected into a booklet which serves as the final report of this grant. The booklets were widely distributed, both to participants in the sessions and also in response to requests by mail.

14. SUBJECT TERMS
Physiology, Psychophysies, Hearing, Auditory

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
			UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

Collected Papers from

the Special Session on

Interactions between

Neurophysiology and Psychoacoustics

Presented at the

117th Meeting of the Acoustical Society of America

Syracuse, New York

May 22, 1989

Session Chairman
and Edited by
Christopher W. Turner

Introduction

The original idea for this special session on Interactions between Neurophsiology and Psychoacoustics came from discussions between Jozef Zwislocki, Robert Smith and myself. With the cooperation of the technical committees and program committee members of Physiological Acoustics, Psychological Acoustics, and Speech Communication, we were able to organize the session as a part of the 117th meeting of the Acoustical Society of America.

There is no doubt that this session could not have taken place without financial support and encouragement by the Air Force Office of Scientific Research. In addition to providing support for the special session itself, the printing costs for this collection of papers were also supported by the AFOSR.

Each of the six invited papers in this collection deals with specific issues of auditory performance. Yet they are tied together by the common theme of relating behavioral performance to the responses of the auditory nervous system which might be used by a subject as a decision variable in the behavioral task. This type of work requires that psychoacousticians, speech researchers, and physiologists work together in using similar paradigms and stimuli, as well as cooperative efforts in the interpretation of the results. It is my hope that these collaborative efforts between auditory physiologists, psychoacousticians, and speech researchers will continue. A place where these groups of auditory researchers can regularly exchange ideas and results will certainly be an asset to the field of auditory research. We thank the Acoustical Society of America for providing such a forum. The interesting papers presented here are an indication that this can be a rewarding enterprise.

Christopher W. Turner
Syracuse University

Table of Contents

Session E Papers

ASA SYRACUSE ABSTRACT

E1. Evolving Ideas of Cochlear Sound Analysis and Stimulus Representation in Hearing. Julius L. Goldstein (Central Institute for the Deaf, St. Louis, MO 63110)

Classical principles of cochlear operation and function in hearing are undergoing major revision because of recent biophysical discoveries that normal cochlear sound analysis is largely determined by centrally controlled nonlinear motor responses from the outer hair cells and that neural temporal entrainment can represent monaural stimulus information. Helmholtz's psychophysically and anatomically based hypothesis of tonotopic cochlear analysis is fully supported, but revision is required of the classical model, he inspired, of that analysis as a tonotopic array of linear filters passively monitored by short-memory energy detectors. Challenges to the classical model have been presented throughout the history of auditory science from psychophysical studies of combination tones, idiotones, masking and periodicity pitch. Hence it is proposed that psychophysics can now be exploited systematically and interactively with biophysical knowledge to contribute to developing the required revised cochlear principles. To get beyond the establishment of *correlations between* psychophysical and biophysical data for specific phenomena, models of some generality are needed for cochlear nonclassical responses and for psychophysical measurement. It is proposed that ideal observer theory provides a general working hypothesis that has been successful in filling the second need. A new signal processing model for nonlinear and active cochlear frequency analysis was formulated to fill the first need. Two modeling studies of the relationship between psychophysics and physiology will be described in detail for periodicity pitch and nonlinear masking.

Author's Popular Version of Paper E1 ASA Syracuse 5/23/89:

Evolving Ideas of Cochlear Sound Analysis and Stimulus Representation in Hearing

J. L. Goldstein, Central Institute for the Deaf, St. Louis, MO, 63110

The inner ear or cochlea is a living organ in two-way communication with the brain that analyzes the world of sound and converts it into patterns of electrical nerve pulses recognized by the brain. The sounds we hear are not uniquely represented in the auditory system, because hearing is probabilistic in nature and the brain must operate very effectively to deal with uncertainty. Until recently, attempts to understand the psychology of hearing in terms of its physiological mechanisms were hampered by lack of awareness that the healthy living cochlea has much finer and more complex powers of analysis than the dead or damaged cochlea. With this barrier removed by improved techniques for studying the cochlea, progress in scientific knowledge of normal and damaged hearing can be accelerated through the integration of interdisciplinary knowledge from psychological, physiological and physical investigations. Progress in understanding how we hear musical notes and how sounds interact in the cochlea exemplify this integration.

The importance of the probabilistic viewpoint of hearing has become clear from psychoacoustical studies of musical note perception. The great 19th century scientist H.L.F. Helmholtz proposed that the cochlea behaves like a harp vibrating sympathetically to sounds rushing past its strings. Different sounds produce distinctive space-time patterns of mechanical vibration upon the harp strings, and the brain is so informed by thousands of sensory hair cells in the cochlea that measure these vibrations and excite patterns of electrical pulses on the auditory nerve. Early this century the success of the low-fidelity telephone system, operating in the frequency range 300-3,000 Hz, made it clear that the low pitch of the human voice and musical sounds could be heard without setting any of Helmholtz's low frequency harp strings below 300 Hz into sympathetic vibration. More recent experiments demonstrated that the brain measures the low pitch from the pattern of vibrating strings in the "cochlear harp" that are excited by overtones remaining in telephone speech. The brain's mechanism for measuring musical notes is unknown, yet many of its properties could be mathematically described by hypothesizing that it uses an optimum scheme for recognizing the low pitch from uncertain measurements of its overtones. Psychoacoustical measurements from people could then be converted into measurements of their internal uncertainty in representing separate overtones within their brain. It was then discovered that an efficient brain measuring the physiological responses to similar sounds from the auditory nerves of laboratory animals can explain what we hear. This approach of assuming the brain is a highly effective processor of uncertain information enables a powerful interaction between psychological and physiological investigations of hearing.

A decade ago it was discovered that the inner ear can emit sounds into the ear canal like a living "magic harp" -- to extend Helmholtz's metaphor. Direct observations of inner ear mechanical vibrations have been studied since early this century, gradually overcoming the nearly insurmountable difficulties of penetrating the toughest bone in the skeleton and recording minute vibrations as small as one hundred millionth of an inch. Recent developments have revealed a highly nonlinear living cochlea, in which three quarters of the ear's hair cells (the "outer" hair cells) participate in determining the mechanical response, while only the remaining "inner" hair cells perform the classical function of passive mechanical sensing. These mechanically active and living outer hair cells provide the healthy ear with an exquisite sensitivity to weak sounds and an acute frequency selectivity, as required by the "cochlear harp". Two-way communication exists between the cochlea and brain. Outer hair cells receive nerve pulses from the brain that modify the ear's mechanical vibrations, while the inner hair cells send electrical pulses to the brain. Many puzzling auditory phenomena that have been known for many years to psychologists and physiologists can now be traced to the "magic harp" and can be exploited to uncover its secrets. A new signal processing theory that attributes these phenomena to nonlinear interactions between living and nonliving vibratory mechanisms is being developed to help understand the underlying biophysics as well as the complex sound analysis functions of the "magic harp".

Invited 40 minute paper for special session "Physiological Acoustics I, Psychological Acoustics I, and Speech Communication I: Interactions between Neurophysiology and Psychophysics" (Sponsored in part by the AFOSR) at the 117th Meeting of The Acoustical Society of America, Syracuse University 22-26 May 1989. Popular version prepared at the request of The American Institute of Physics, Public Information Division.

ASA Syracuse 23 May 1989 Session E: Interactions between Neurophysiology & Psychoacoustics

E1: Evolving Ideas of Cochlear Sound Analysis and Stimulus Representation in Hearing. J.L. Goldstein, Central Inst. for the Deaf, St Louis, MO 63110

TEXT OF LECTURE

(Thanks to session chairman, Chris Turner, for the invitation to open this session with an historical overview along with discussions of my own work.)

Slide 1: Outline

I. The first part of my talk outlines the historical evolution of key ideas on cochlear sound analysis and stimulus representation in hearing. Three stages of development are considered corresponding approximately to the states of knowledge in the years 1857, 1966 and 1986.

To streamline the historical review I will mention mainly physical evidence concerning cochlear operation. Of course history is replete with inferences of cochlear operation based upon psychoacoustics.

II. The second part of my talk will compensate for this with two detailed examples of strong interactions between physiology and psychoacoustics.

Slide 2: Helmholtz's Harp

We begin with Helmholtz's hypothesis of the Organ of Corti as a spatial array of resonators driving the auditory nerve. Adopting the popular metaphor of his day, I refer to this frequency-place transform as "Helmholtz's Harp". The pictures at the top of the slide are from Helmholtz's 1857 popular lecture. The resonator frequency response is from his Sensations of Tone. The modern functional view of the cochlea as filter-bank spectrum analyzer is just a paraphrase of "Helmholtz's Harp".

We see that Helmholtz investigated the perception of complex sounds from speech and music. His physiological hypothesis concerning cochlear frequency analysis was an inference based upon psychoacoustical data on the perception of vowel formants and the harmonic structure of musical sounds.

Slide 3: The Inert Ear & Probabilistic Nature of Hearing

100 years later we have better stimulus control (Weiner & Ross) with a focus on simple stimuli in psychoacoustical and physiological studies.

Bekesy's (1942, 1960) measurements and theoretical suggestions have simplified cochlear frequency analysis to inert basilar membrane hydromechanics.

The probabilistic nature of hearing has been discovered in the psychoacoustics of simple tasks, and in auditory-nerve spike patterns. This is shown by the ROC data and theory for signal detection developed by Green & Swets, and in the Poisson-like properties of auditory-nerve spike discharges measured by Kiang and colleagues. These findings suggest that the internal representation of stimulus information is best described in terms of random decision variables, that can be efficiently processed by the brain.

Zwislocki (1950) showed that hydromechanical theory was capable of accounting for Bekesy's observations. This work established a productive branch of biophysics.

The discrepancy between auditory-nerve and basilar membrane tuning could be reconciled by postulating an intermediate mechanism that responds to a spatial derivative of basilar membrane displacement (Bekesy, 1953).

Slide 4: The Vital Ear & Nonlinear Nature of Cochlear Analysis

About 20 years have passed. We now have a focus on studies of responses to complex stimuli in psychoacoustics and physiology (e.g. Young & Sachs).

The mechanism and properties of cochlear frequency analysis are far more complex than previously believed (Lim, 1986). The OHC's, which are under efferent control (Mountain, 1980; Siegel & Kim, 1982), have been discovered to be responsible for a highly sensitive, selective and compressive tuned basilar membrane response at all but the most intense sound levels (Rhode, 1971; Sellick, et al, 1982; Robles, et al, 1986). In brief, cochlear mechanics is a vital process.

In psychoacoustical modeling for complex stimuli there is increased interest in probabilistic psychoacoustic experiments (e.g. Miller & Nicely, Houtsma & Goldstein).

The cochlea is still thought of as a frequency analyzer, and it has been shown (Sachs & Young, 1980) that this function can be performed best if auditory-nerve spike timing, rather rate alone, is processed by the brain.

Otoacoustic emissions have been discovered as an exciting byproduct of the vital cochlear mechanics (Kemp, 1978; Wilson, 1980). This suggests that an appropriate popular metaphor for the cochlea can be the "magic harp".

Slide 5: Periodicity-Pitch Psychophysics

Periodicity-Pitch psychophysics is our first example of the interaction between psychoacoustics and neurophysiology.

Houtsma and I found that the fundamental period of a sound can be heard when its overtones are presented to different ears, thereby proving the existence of a central pitch mechanism.

Slide 5 shows results for psychoacoustical experiments with periodic sounds consisting of two randomized successive harmonics. Probabilities for correct identification of eight neighboring notes on the musical scale were measured with musically experienced subjects. The data, plotted as equal probability contours, show the effects of harmonic numbers of the overtones and the reference fundamental (the first note). A similar advantage was found for lower harmonics whether they were presented to one ear (monotic) or split between the two ears (dichotic).

The dichotic experiments proved that a central periodicity-pitch mechanism exists that is capable of integrating neural signals representing pure tones. Since the monotic and dichotic results are similar, we concluded that cochlear frequency analysis provides the central pitch mechanism with similar signals in both cases.

Slide 6: Psychophysical Theory of Fo Perception

A probabilistic psychophysical theory of periodicity-pitch for complex tones was formulated that postulates the existence of several internal random decision variables, one for each analyzed stimulus harmonic. Harmonic frequency information is carried by each decision variable. The standard deviations of the Gaussian decision variables are the free parameters of the model.

A central pitch processor is postulated that provides an optimum estimate of the fundamental frequency corresponding to the noisy harmonic frequency measurements. Using this model, the standard errors of the decision variables were calculated from the probabilistic psychophysical data as shown next in slide 7.

(Skip template)

Slide 7: Fo Theory & Data

Calculated standard errors of the model decision variables are shown for three subjects from our psychophysical study. The minima range from 0.6% to 1%. All three subjects show a similar sharp deterioration at harmonic frequencies above 2 to 3 kHz. This property suggested the possibility that the decision variable represents a neural timing signal. (Skip the rest.)

Slide 8: Siebert 1968, 1970

Siebert did pioneering research on applying optimum statistical processing to account for probabilistic psychoacoustics with probabilistic auditory-nerve discharges. I applied his work to study the relation between the optimum pitch processor decision variables and his model of auditory nerve statistics. This classroom slide shows an overview of Siebert's work that I taught to my students for over 15 years.

Slide 9: Neural Representation of Component Frequencies

This slide summarizes theoretical work done with Peter Srulovicz on the neural basis of the decision variable in the optimum processor for periodicity-pitch. Comparisons between neural and psychophysical standard errors are shown at the lower left.

Curve C is the expected standard error for optimum frequency measurements on neural interspike intervals from a single auditory-nerve fiber tuned to the frequency being measured. Curve A is the average (from slide 7) for the psychophysical decision variable in the optimum periodicity-pitch processor. Curve B is Moore's psychophysical measurement of pure tone frequency discrimination. All three curves are similarly dependent upon frequency.

The top figure illustrates schematically that cochlear frequency analysis provides robust temporal information on component frequencies in a complex stimulus. Nerve fibers that are tuned to a component tone generally provide the best temporal representation for that tone.

Our model for measuring component frequencies of a complex tone from tuned neural synchrony responses is shown at the lower right. The matched filter assures that only interspike intervals closely matching the CF are measured. This "tuned synchrony spectrum analyzer" reflects the statistics of auditory-nerve fibers and accounts for the psychophysical decision variables in periodicity-pitch perception.

Slide 10: Neural Representation of Voice Pitch

Miller and Sachs provided a direct neurophysiological demonstration that tuned synchrony responses from the auditory nerve are the most effective signals for mediating periodicity-pitch. They examined large populations of auditory-nerve fiber responses to voiced speech sounds, in quiet and in background noise.

The top two figures show their ALSR spectra for a 25 ms. vowel response segment, in quiet and in noise. The ALSR is Young and Sachs' algorithm for measuring the tuned synchrony spectrum from sampled population response data. Strong spectral peaks are found at low harmonic frequencies of the 120 Hz fundamental in both cases. Calculations with the cepstrum algorithm (middle) show that the fundamental period is robustly represented by the ALSR spectra.

The fundamental period can also be measured directly from responses of individual auditory-nerve fibers. This is shown, at the bottom, as a function of fiber CF for the same population data. Fibers tuned to the fundamental energy in the stimulus provide Fo information in both cases. Fibers tuned to beating high harmonics provide Fo information in quiet only.

From psychoacoustics it is known that Fo is perceived for the vowel in noise and in the absence of fundamental stimulus energy. Mediation of periodicity-pitch by tuned synchrony responses from low stimulus harmonics is clearly supported.

Summary of first example.

The psychophysical and neurophysiological studies of periodicity-pitch provide powerful evidence for the existence of a central processor that efficiently integrates tuned synchronized responses from a broad range of different auditory-nerve fibers. From psychoacoustics it is known that this processor is the dominant mechanism for measuring periodicity pitch. The stimulus information processed can be represented as tuned synchrony spectra. However, detailed knowledge of the actual algorithm or neuralware used by the central processor requires neurophysiological study beyond the auditory nerve. Optimum probablistic models of central processing avoid being prematurely specific on processor implementation.

Slide 11: Nonlinear Cochlear Sound Analysis

Nonlinear cochlear sound analysis is my second example of the interaction between neurophysiology (top) and psychoacoustics (bottom).

Auditory nerve and psychoacoustical studies of two-tone suppression and combination tones are highly correlated. The earliest systematic auditory nerve studies involved a close interaction with psychoacoustics. Sachs and Kiang's 2TS study stimulated Houtgast's (1972) demonstration of psychoacoustical suppression in forward masking. At the lower left are data from Duifhuis' (1980) psychophysical study of 2TS, which is closely correlated with Abbas and Sachs' (1976) auditory-nerve study, both demonstrating a strong level-dependent effect of low frequency suppressors.

The interaction on combination tones, shown at the right, progressed from psychoacoustics to the auditory nerve. The earliest systematic psychoacoustical evidence or nonlinear cochlear analysis was given in 1924 by Wegel and Lane (bottom center), with their discovery of the nonlinear growth of masking.

An example (center), recently presented by Kiang, illustrates both phenomena in responses from a single auditory-nerve fiber. A and B show threshold tuning curves measured in quiet and with a constant background tone (65 dB SPL @ 7 kHz). The presence of the constant tone causes an 8 dB suppression at the tuning curve tip, and the addition of an intermodulation response lobe.

These correlated demonstrations of nonlinear cochlear analysis contributed to the reevaluation of accepted understanding of intracochlear mechanisms during the past 20 years, leading to current knowledge of the vital cochlear and nonlinear basilar membrane mechanics. This, later, intracochlear knowledge has proved necessary for progress on comprehensive models of data on nonlinear cochlear sound analysis from psychoacoustics and the auditory nerve.

Slide 12: Multiple Bandpass Nonlinearity Filter Model

My current work on a signal processing model for basilar membrane mechanics is shown in slide 12. The MBPNL dual filter model (top) represents the nonlinear mechanical transformation between stapes input and basilar membrane output. The sensitive "tip" of the basilar membrane mechanical tuning curve is now known to be compressive at above threshold sound levels (>15-35 dB SL). It is known from early work on CTs and 2TS that a compressive BPNL filter can represent considerable data. Therefore a BPNL filter is chosen in the lower transmission path to represent the tip response. An amplifier under MOC efferent control is included, following Gifford and Guinan (1983).

Basilar membrane measurements with stimulus tones in the low frequency "tail" region of the mechanical tuning curve demonstrate linear-like excitatory responses along with nonlinear compression of tip transmission. These properties are provided by the upper MBPNL transmission path, which interacts with the lower path. The inverse nonlinear transducers provide the required linear throughput for tail tones.

Simulated MBPNL filter responses for simple tone stimuli are shown at the lower right. The top curve represents a normal CF tone response. At low levels it is linear, with a threshold response of about 3 Angstroms. Beyond 25 dB it is compressive; the 0.5 dB/dB compression shown is typical for a filter CF of 1 kHz. AT 95 dB the tip and tail responses interfere, to produce a sharp response notch. At still higher levels the linear tail response dominates being about 1 micron at 110 dB SPL; this is where Bekesy performed his measurements.

The horizontal separation between the two component responses at low sound levels is the tip-tail separation (about 40 dB). Reduction of the tip gain by 20 dB produces the lower notched response. Elimination of the tip gain produces the linear "inert" cochlea. Addition of a second signal (in this case DC) suppresses the CF response without modifying the notch.

Sound level dependence, for broadband stimuli, is the outstanding functional property of the MBPNL filter. It responds like a bandpass filter at low sound levels, and like a lowpass filter at high sound levels. (This should be clear from the composite tuning curve at the lower left.) The nonlinear interaction between the two filters sharpens the level-dependent transition between these two states. This property is considered next.

Slide 13: Auditory Nerve Data for Sound Level Dependence of Cochlear Frequency Analysis

Sachs and colleagues measured synchronized complex tone responses from auditory nerve fibers as a function of absolute sound level. Here are their data from a fiber stimulated with a 1.90 kHz CF tone plus a 1.14 kHz low frequency tone (x axis). The lower tone is always 15 dB more intense. They measured Fourier coefficients vs. sound level for period histogram responses to the two-tone stimulus.

The CF response dominates at low levels (indicate regions), followed by a transition region at intermediate levels, and low frequency dominance at the higher levels. Note that the low level CF response is similar to the response of the CF tone alone.

These data illustrate the transition from bandpass to lowpass filter characteristic with increasing sound stimulus level. This behavior can be modeled with the MBPNL filter, as shown next.

Slice 14: MBPNL Simulation of Sachs, et al (1980)

The Sachs, et al data are simulated with the same model parameters as used before in Slide 12. The MBPNL response to the CF tone alone is the same as before. To simulate the two-tone response, the low frequency tone is attenuated 30 dB by the tip bandpass filter and not at all by the tail lowpass filter, producing the two notched responses shown. Notches are not apparent in the neural data. A quadrature (90 deg) phase relation between the two transmission paths is more consistent with the data (as simulated at the upper right).

Data and theory are compared (bottom) by considering the ratio of the Fourier coefficients for the neural data and the MBPNL (quadrature) simulation. The ratio measure factors out, to a first approximation, the mechanical to neural transformation of

the auditory nerve data. The log ratio of the low-to-high frequency response is plotted; it is about -15 dB for low sound levels and about +15 dB for high sound levels. Data and theory are in good agreement, showing a similar transition. (The neural data have been shifted 23 dB to the right to conform to the lower absolute sensitivity of the "human" model.)

Slide 15: Relation Between Psychoacoustical Masking and Nonlinear Cochlear Mechanics (MBPNL viewpoint)

The MBPNL model describes nonlinear aspects of psychophysical masking that previously were problematic. This is next briefly considered in slide 15.

Shown at the top are rates of masking growth for simultaneous masking measured by Wegel and Lane (1924) with an 800 Hz masker, along with similar contemporary data. For target frequencies larger than about twice the masker frequency, the data are consistent with a dominance of tail transmission of the masker tone and tip transmission of the target tone. From more detailed data on simultaneous masking it is also possible to distinguish between suppressive and excitatory masking by low frequency tones.

Similar dual path transmission is reflected in Duifhuis' (1980) data on psychophysical suppression, which are summarized at the bottom. The steep rates of suppression found for low frequency suppressors is similarly explained by the dual transmission paths taken by the stimulus. The strong psychophysical suppression results from the nonlinear interaction between the two paths.

Summary of second example.

I have proposed that the MBPNL signal processing model, which represents observations on nonlinear basilar membrane mechanics, provides a new basis for understanding data from the auditory nerve and psychoacoustics.

Slide 16: Schouten's (1970) Research Triangle

I wish to end my talk by recalling Schouten's humorous caricature of the mutual interactions between psychophysics and physiology that is the subject of this session. Schouten describes the mutual interactions as eternal and includes theory as a separate third enterprise. The triangle of enterprises with traffic in both directions were all exemplified in the two detailed examples I discussed. The eternal triangle was also present in the actual research I overviewed in the first part of my talk. I mentioned that Helmholtz's hypothesis concerning cochlear operation was a synthesis of psychoacoustics and anatomy. Of course it was mediated by Fourier theory and Ohm's law of hearing. What is truly different in contemporary research is that

rapid developments in neurophysiology and biophysics demand more responsive and efficient mutual interaction. This requires an acceptance of theory as a partner in the research triangle and a prevailing good humor needed for psychoacousticians, physiologists and theoreticians to interact aimicably. This I think was Schouten's message.

REFERENCES

Abbas, P.J. and Sachs, M.B. (1976). Two-tone suppression in auditory-nerve fibers: Extension of stimulus response relationship. J. Acoust. Soc. Am., 59, 112-122.

Bekesy, G. von, (1960). Experiments in Hearing. (E.G. Wever, ed.) McGraw-Hill, NY.

Brödel, M. (1946). Three unpublished drawings of the anatomy of the human ear. Saunders, Phil., PA.

Duifhuis, D.H. (1976). Cochlear nonlinearity and second filter: Possible mechanism and implications. J. Acoust. Soc. Am., 59, 408-423.

Duifhuis, D.H. (1980). Level effects in psychophysical two-tone suppression. J. Acoust. Soc. Am., 67, 914-927.

Gifford, M.L. and Guinan, J.J. (1983). Effects of crossed -olivocochlear- bundle stimulation on cat auditory nerve fiber responses to tones. J. Acoust. Soc. Am., 74, 115-123.

Goldstein, J.L. (1967). Auditory nonlinearity. J. Acoust. Soc. Am., 41, 676-689.

Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am., 54, 1496-1516.

Goldstein, J.L. (1978). Mechanisms of signal analysis and pattern perception in periodicity pitch. Audiology, 17, 421-445.

Goldstein, J.L. (1989). Updating cochlear driven models of auditory perception: A new model for nonlinear auditory frequency analysing filters. In Working Models of Human Perception, B. Elsendoorn and H. Bouma, eds., Academic Press, pp. 19-57.

Goldstein, J.L. and Kiang, N.Y.S. (1968). Neural correlates of the aural combination tone $2f_1-f_2$. Proc. IEEE, 56, 981-992.

Goldstein, J.L. and Srulovicz, P. (1977). Auditory nerve spike intervals as an adequate basis for aural frequency measurement. In E.F. Evans and J.P. Wilson (eds.) Psychoacoustics and Physiology of Hearing. (Academic, London), pp. 337-346.

Goldstein, J.L., Gerson, A., Srulovicz, P. and Furst, M. (1978). Verification of the optimum probabilistic basis of aural processing in pitch of complex tones. J. Acoust. Soc. Am., 62, 486-497.

Green, D.M. and Swets, J.A. (1966). Signal Detection Theory and Psychophysics. (Wiley, NY).

Helmholtz, H. von, (1857). On the physiological causes of harmony in music. In Helmholtz's Popular Scientific Lectures, M. Kline, ed. (Dover, NY, 1962) pp, 22-58.

Helmholtz, H.L.F. (1863). Sensations of Tone. (Dover, NY, reissued 1954).

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Am., 51, 1885-1894.

Houtsma, A.J.M. and Goldstein, J.L. (1972). The central origin of the pitch of complex tones: evidence from musical interval recognition. J. Acoust. Soc. Am., 51, 520-529.

Huggins, W.H. and Licklider, J.C.R. (1951). Place mechanisms of auditory frequency analysis. J. Acoust. Soc. Am., 23, 290-299.

Kemp, D.T. (1978). Stimulated acoustic emissions from within the human auditory system. J. Acoust. Soc. Am., 64, 1386-1391.

Kiang, N.Y.S. (1984). Peripheral neural processing of auditory information. In Handbook of Physiology-The Nervous System III, I. Darian-Smith, ed., ch. 15, 639-674. The American Physiological Society, Bethesda, MD.

Kiang, N.Y.S., et al (1965). Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. (MIT Press, Cambridge.)

Lim, D.J. (1986). Functional structure of the organ of Corti: a review. Hearing Res., 22, 117-146.

Miller, G.A. and Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am., 27, 338-352.

Miller, M.I. and Sachs, M.B. (1984). Representation of voice pitch in discharge patterns of auditory nerve fibers. Hearing Res., 14, 257-279.

Mountain, D.C. (1980). Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics. Scinece, 210, 71-72.

Ohm, G.S. (1843). On the definition of tones. Ann. Phys. Chem., 59, 513-565.

Rhode, W.S. (1971). Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique. J. Acoust. Soc. Am., 49, 1218-1231.

Robles, L., Ruggero, M.A. and Rich, N.C. (1986). Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves and response phases. J. Acoust. Soc. Am., 80, 1364-1374.

Sachs, M.B. and Kiang, N.Y.S. (1968). Two-tone inhibition in auditory-nerve fibers. J. Acoust. Soc. Am., 43, 1120-1128.

Sachs, M.B. and Young, E.D. (1980). Effects of nonlinearities on speech encoding in the auditory nerve. J. Acoust. Soc. Am., 68, 858-875.

Sachs, M.b., Young, E.D., Schalk, T.B. and Bernardin, C.P. (1980). In G. van den Brink and F. A. Bilsen, (1980), Psychophysical, Physiological and Behavioural Studies in Hearing. (Delft Univ. Press, Delft) pp. 284-292.

Schouten, J.F. (1970). The residue revisited. In R. Plomp and G.F. Smoorenburg, eds. Frequency Analysis and Periodicity Detection in Hearing. (Sijthoff, Leiden) pp. 41-54.

Sellick, P.M., Patuzzi, R. and Johnstone, B.M. (1982). Measurement of basilar membrane motion in the Guinea pig using the Mossbauer technique. J. Acoust. Soc. Am., 72, 131-141.

Siebert, W.M. (1970). Frequency discrimination in the auditory system: place or periodicity mechanisms? Proc. IEEE, 58, 723-730.

Siegel, J.H. and Kim, D.O. (1982). Efferent neural control of cochlear mechanics? Olivocochlear bundle stimulation affects cochlear biomechanical nonlinearlity. Hearing Res., 6, 171-182.

Srulovicz, P. and Goldstein, J.L. (1983). The central spectrum: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. J. Acoust. Soc. Am., 73, 1266-1276.

Wegel, R.L. and Lane, C.E. (1924). The auditory masking of one pure tone by another and its possible relation to the dynamics of the inner ear. Phys. Rev., 23, 226-285.

Wiener, F.M. and Ross, D.A. (1946). The pressure distribution in the auditory canal in a progressive sound field. J. Acoust. Soc. Am., 18, 401-408.

Wilson, J.P. (1980). Evidence for a cochlear origin for acoustic re-emission, threshold fine-structure, and tonal tinnitus. Hearing Res., 2, 233-252.

Young, E.D. and Sachs, M.B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns Oxof populations of auditory-nerve fibers. J. Acoust. Soc. Am., 66, 1381-1403.

Zwislocki, J. (1950). Theory of the acoustical action of the cochlea. J. Acoust. Soc. Am., 22, 778-784.

Evolving Ideas of Cochlear Sound Analysis and Stimulus Representation in Hearing

J.L. Goldstein, Central Inst.for the Deaf, St. Louis, MO 63110

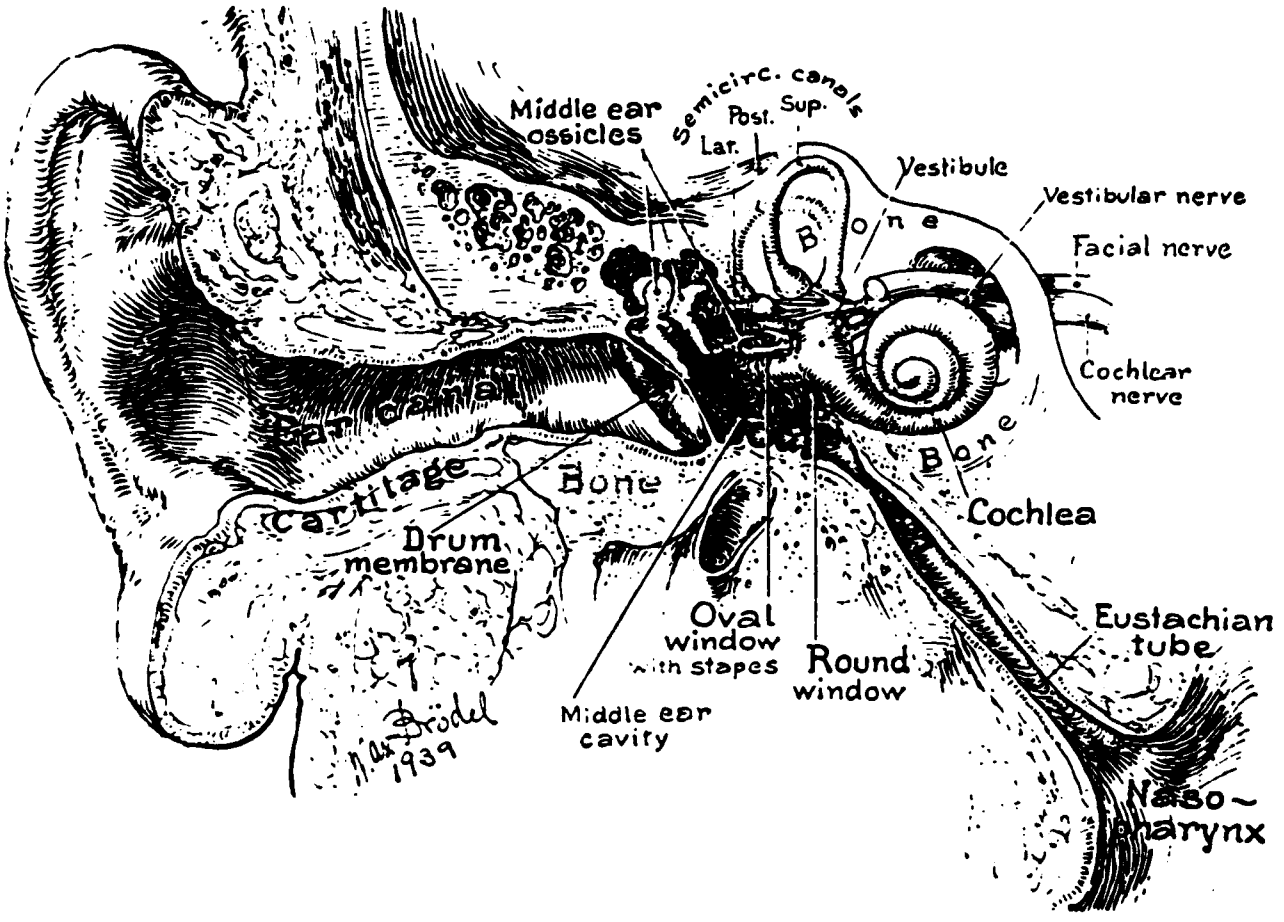


Fig. 1.2. Reproduction of M. Brödel's classical drawing of the cross section of the human ear. (From Brödel, 1946.)

Outline of Talk

I. Evolution of Ideas
   1) 1857: Helmholtz's Harp
   2) 1966: The Inert Cochlea & Probabilistic Hearing
   3) 1986: The Vital Cochlea & Nonlinear Cochlear Analysis

II. Quantitative Relations between Physiology & Psychoacoustics
   1) Periodicity Pitch & Optimum Probabilistic Processing
   2) Cochlear Nonlinear Analysis & Psychophysical Masking

HELMHOLTZ'S HARP (1857, 1863)

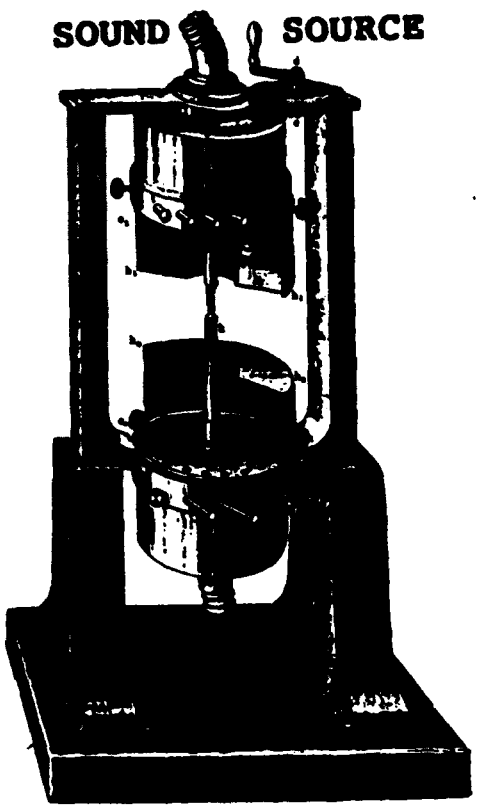
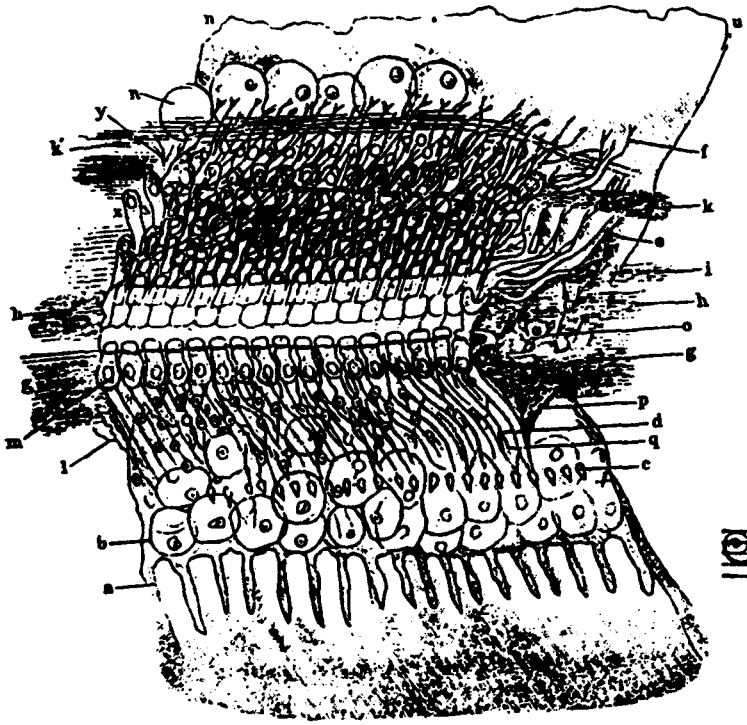
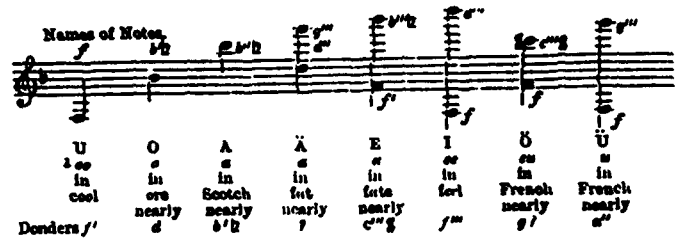


FIG. 1

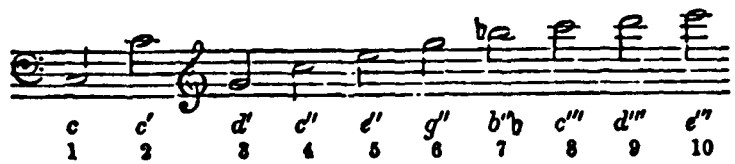


ORGAN OF CORTI

PERCEPTS

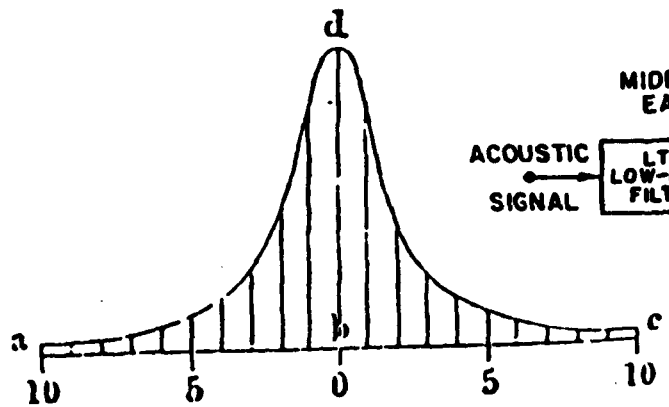


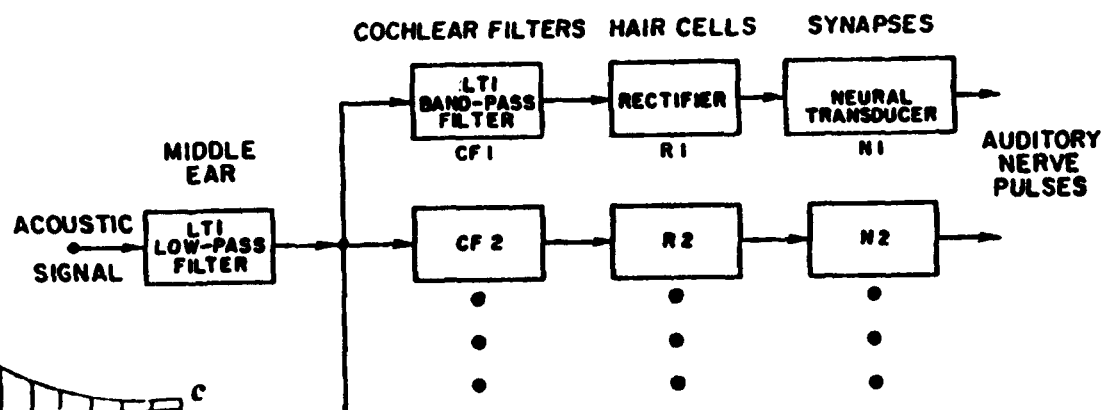
VOWELS



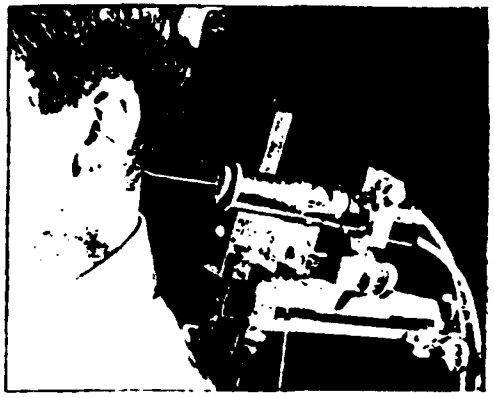
MUSIC

FIG. 52.

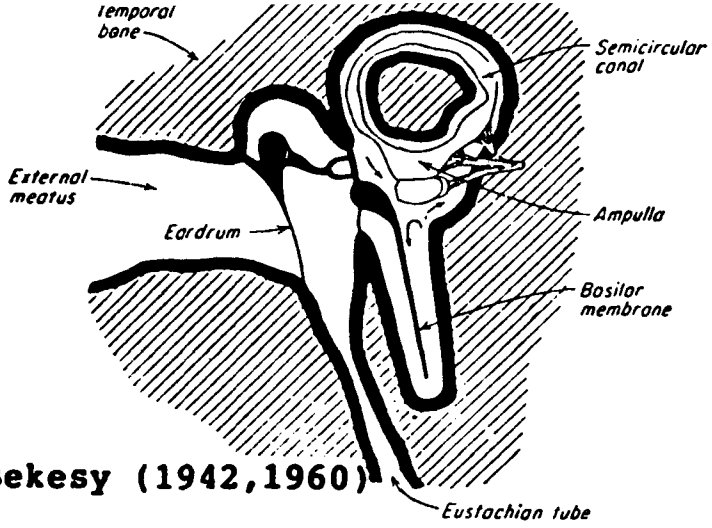




Filter-Bank Integrative Cochlear Model



Wiener & Ross (1946)



Bekesy (1942,1960)

FIG. 11-24. Schematic section of the ear.

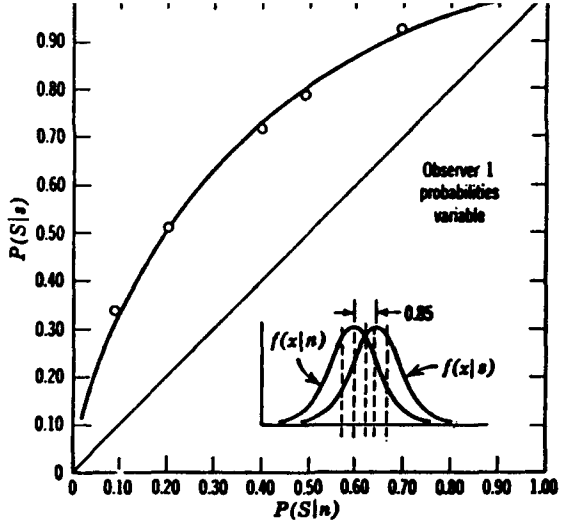
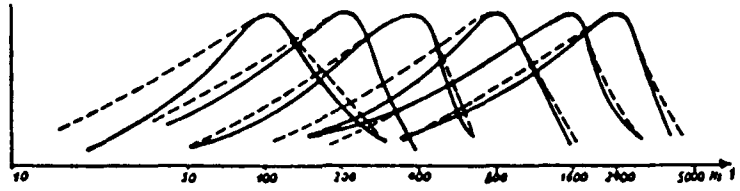
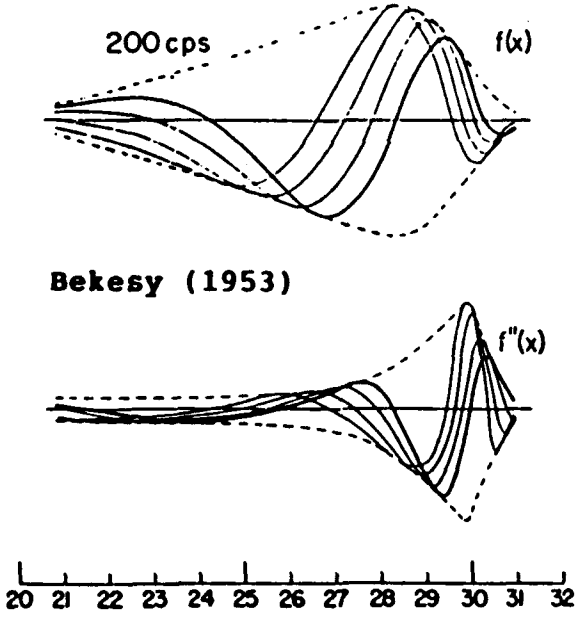


FIG. 4-1 An empirical ROC graph and a theo-

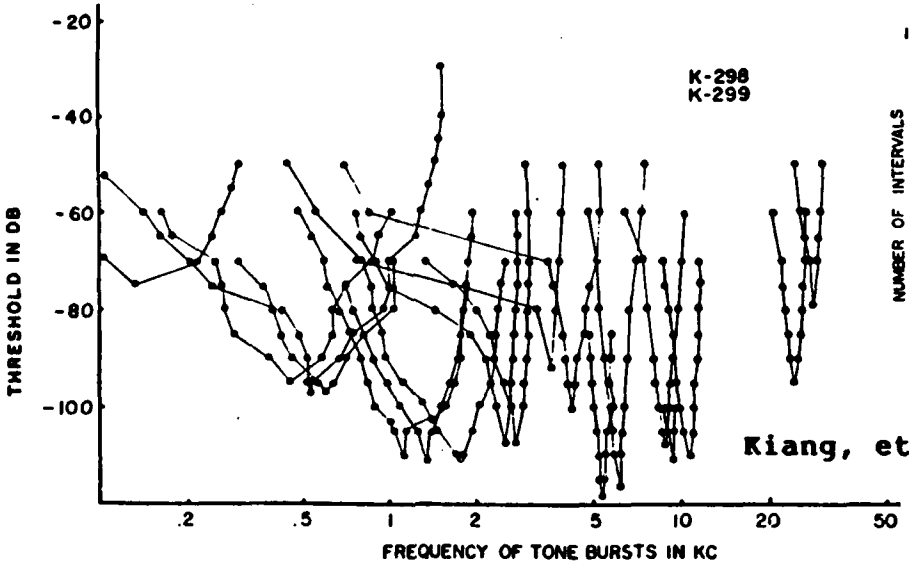
Green & Swets (1966)

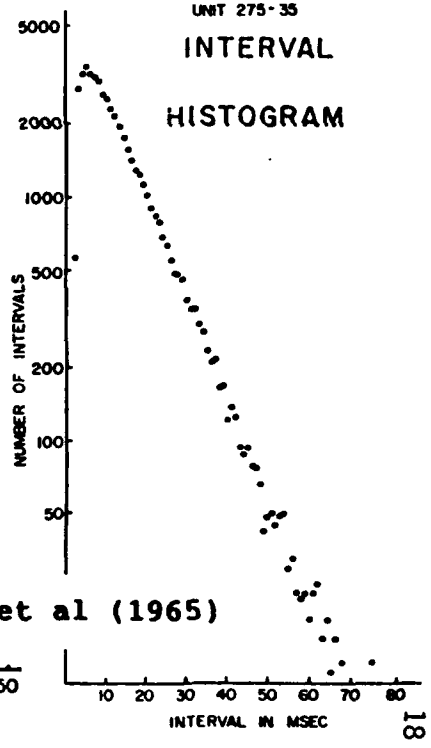
Zwislocki (1950)





Bekesy (1953)

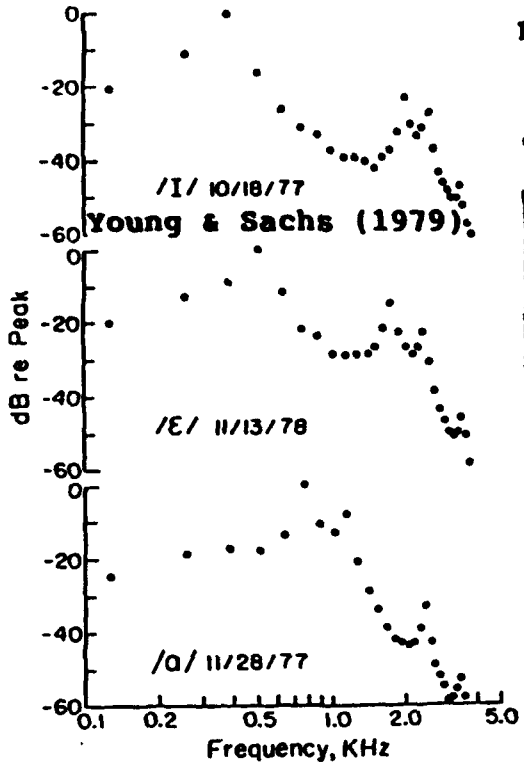




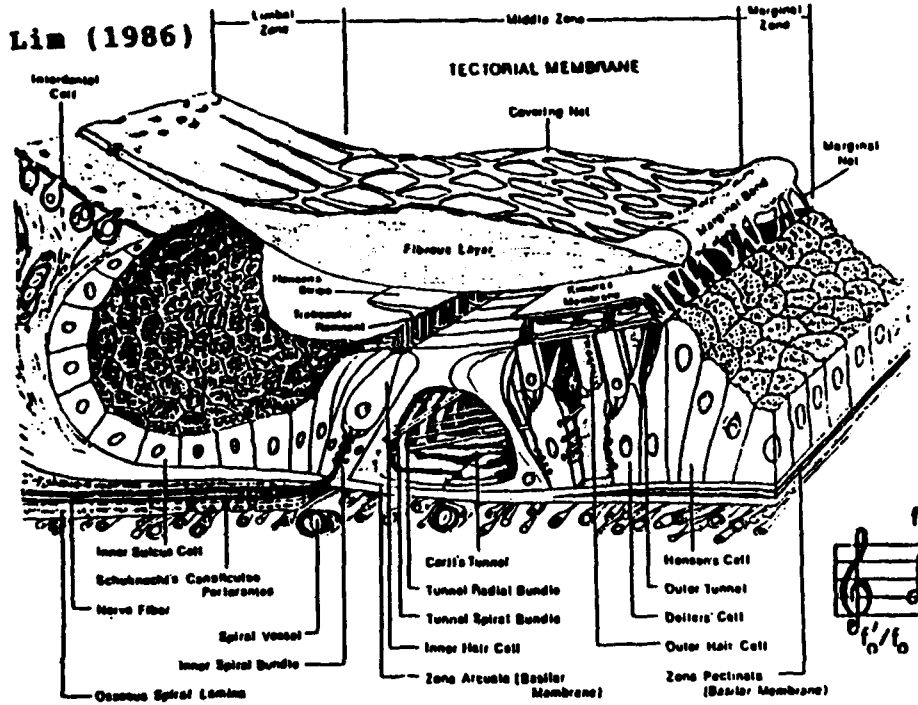
Kiang, et al (1965)

THE VITAL EAR & NONLINEAR NATURE OF COCHLEAR ANALYSIS

Young & Sachs (1979)



Lim (1986)

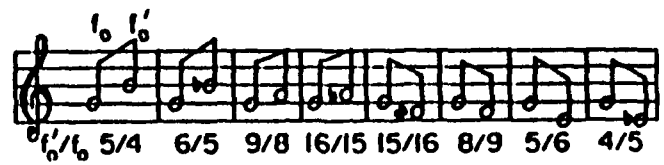


An artist's conception of the organ of Corti. (Drawing by Nancy Sally.)

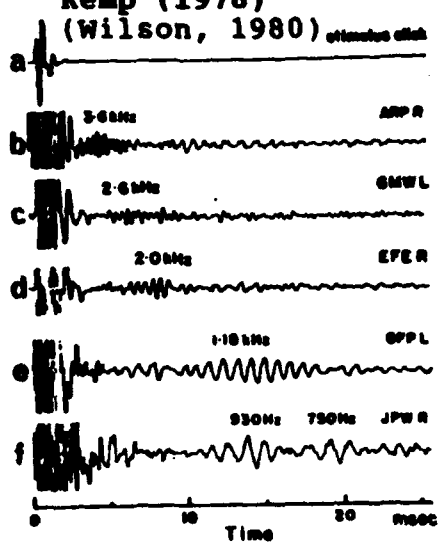
Miller & Nicely (1955)

comparison of 16 consonants:
$|p|$, $|t|$, $|k|$, $|f|$, $|\theta|$,
$|s|$, $|ʃ|$, $|b|$, $|d|$, $|g|$,
$|v|$, $|ð|$, $|z|$, $|ʒ|$, $|m|$,
and $|n|$.

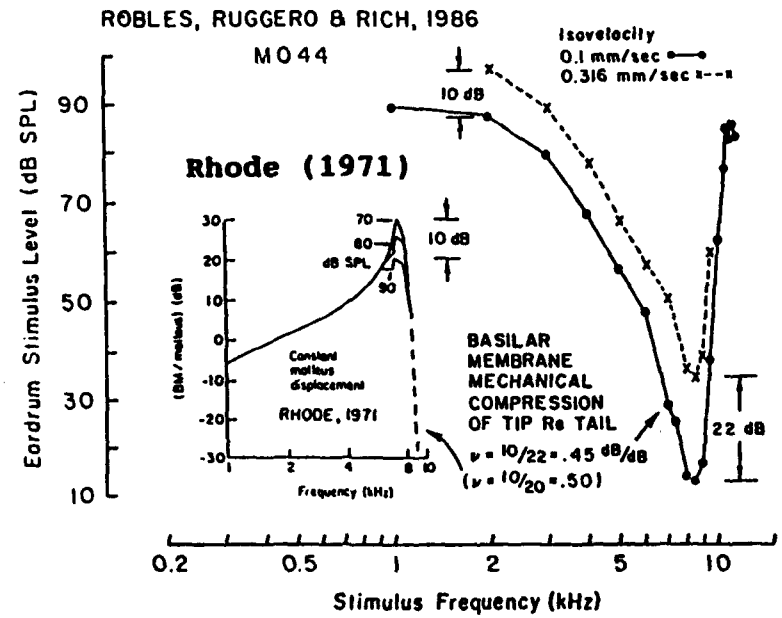
Houtsma & Goldstein (1972)



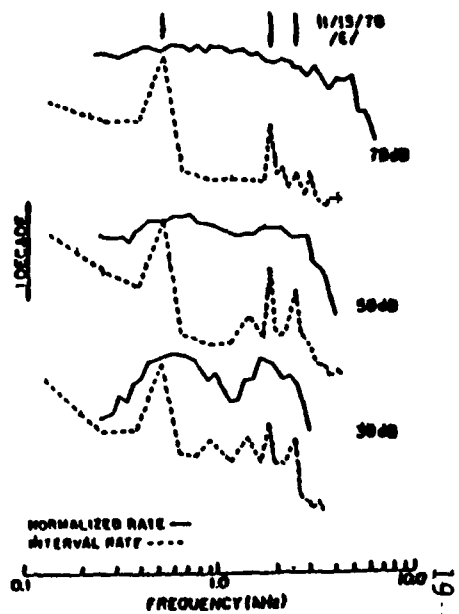
Kemp (1978)
(Wilson, 1980)



ROBLES, RUGGERO & RICH, 1986

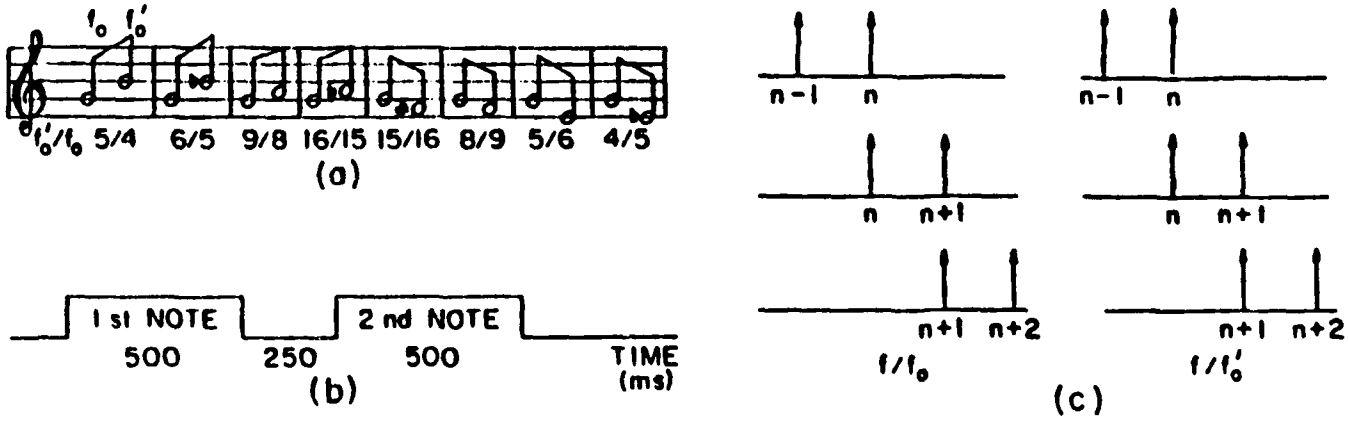


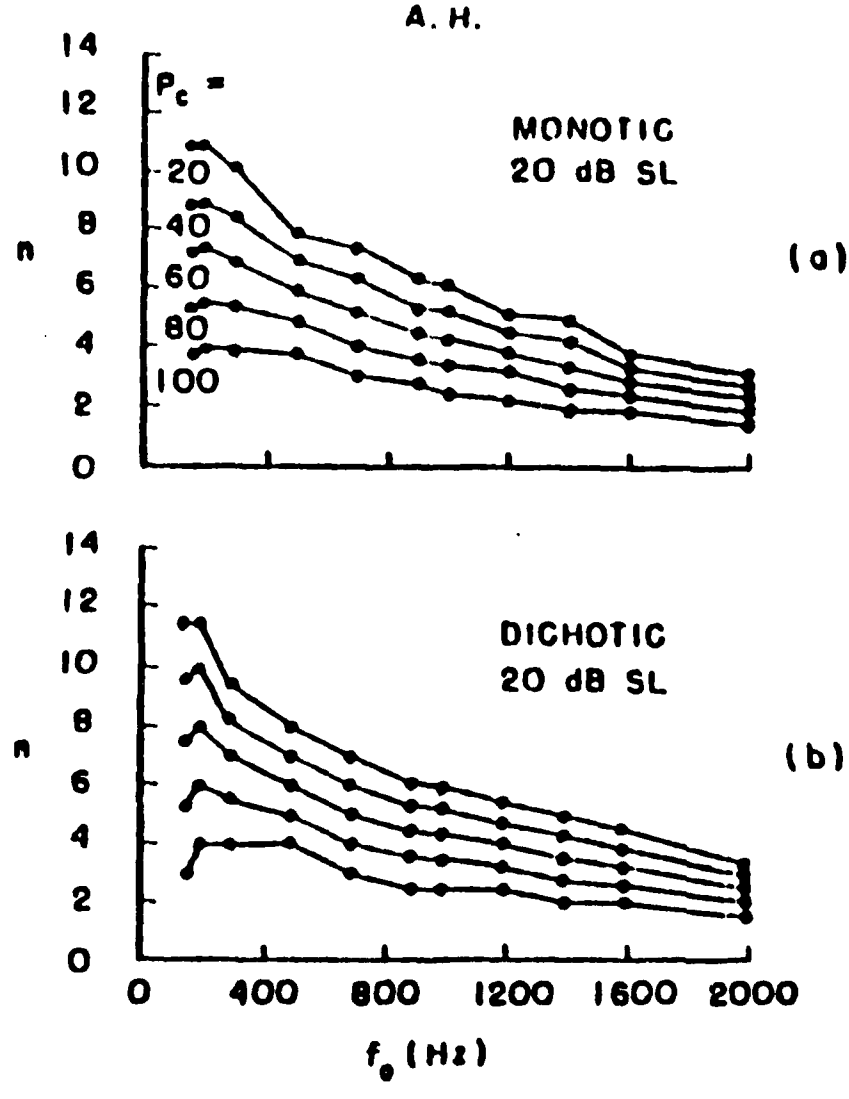
Sachs & Young (1980)



PERIODICITY-PITCH PSYCHOPHYSICS

Houtsma & Goldstein (1972)

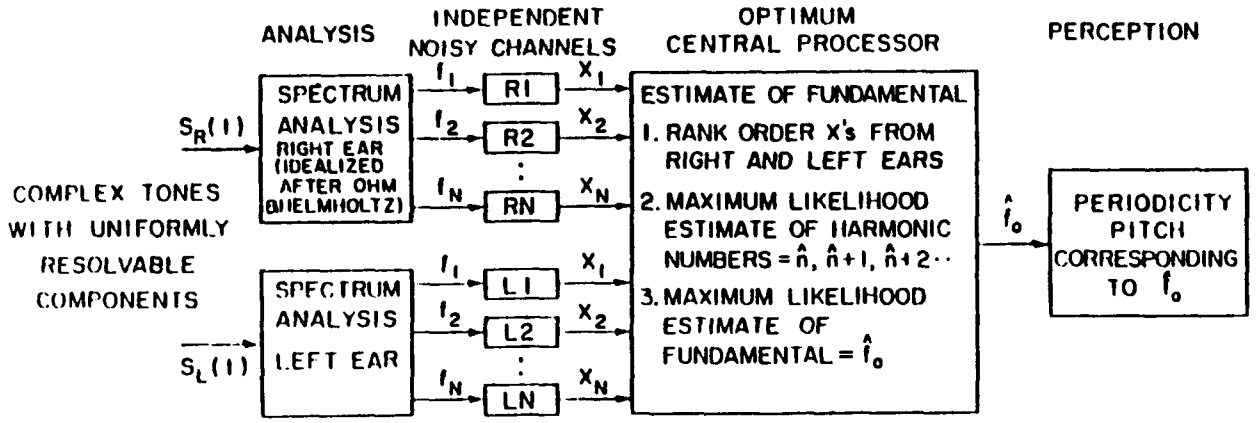


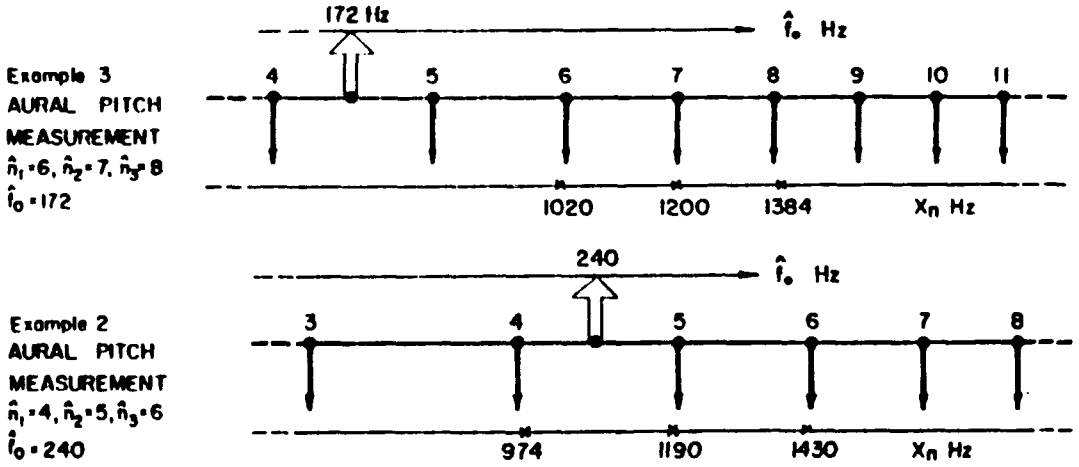


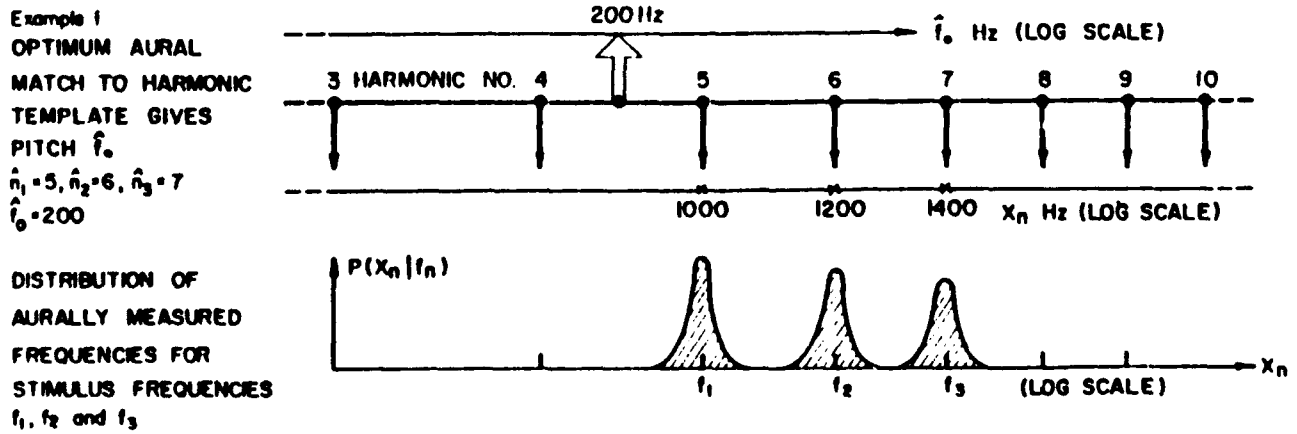
slide 6

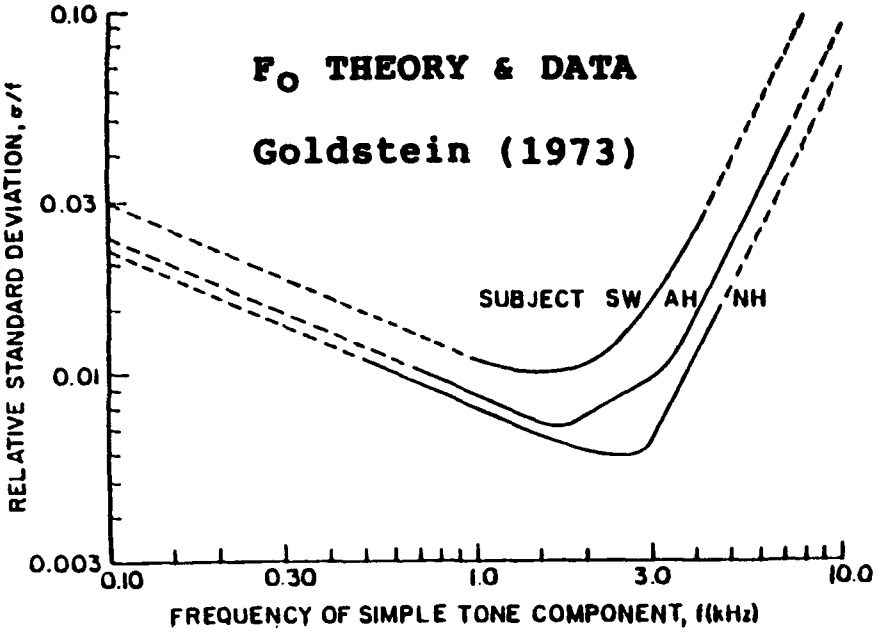
PSYCHOPHYSICAL THEORY OF F_0 PERCEPTION

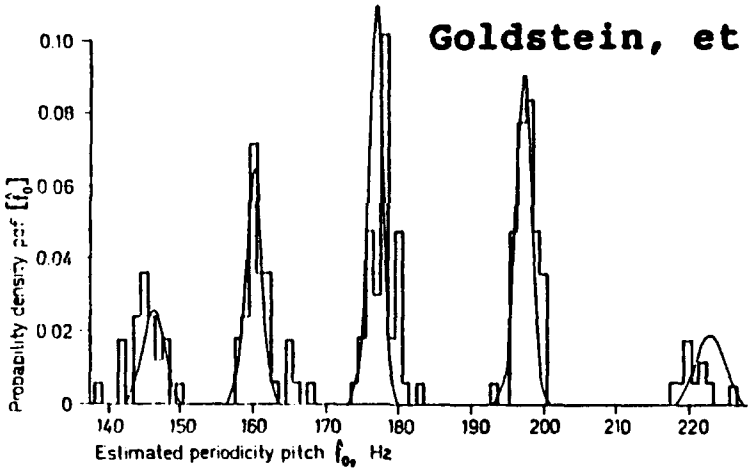
Goldstein (1973)

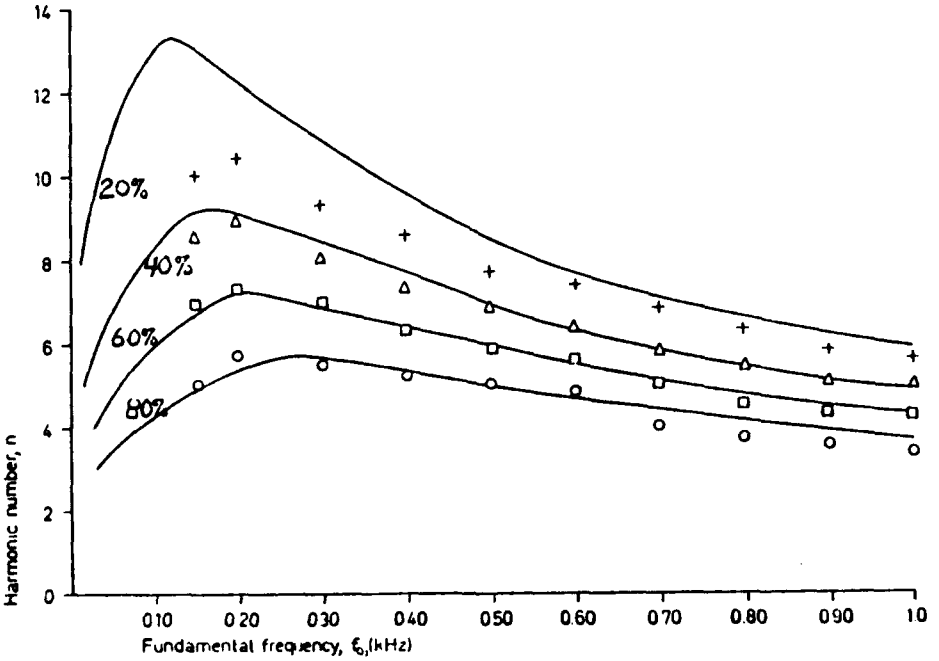







F_o THEORY & DATA
Goldstein (1973)
SUBJECT SW AH NH
RELATIVE STANDARD DEVIATION, σ/f
0.10
0.03
0.01
0.003
0.10
0.30
1.0
3.0
10.0
FREQUENCY OF SIMPLE TONE COMPONENT, f(kHz)


Goldstein, et al (1978)
Probability density pdf [f0]
0.10
0.08
0.06
0.04
0.02
0
140
150
160
170
180
190
200
210
220
Estimated periodicity pitch f0, Hz


20%
40%
60%
80%
Harmonic number, n
14
12
10
8
6
4
2
0
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.0
Fundamental frequency, f0,(kHz)

SIEBERT 1968, 1970

OPTIMUM PROCESSING OF AUD. NV. SPIKES: FREQ. DISCRIM. OF SIMPLE TONE (1 kHz)

$$d'_{cF} = \beta \frac{\Delta f}{f}$$

Model Parameters

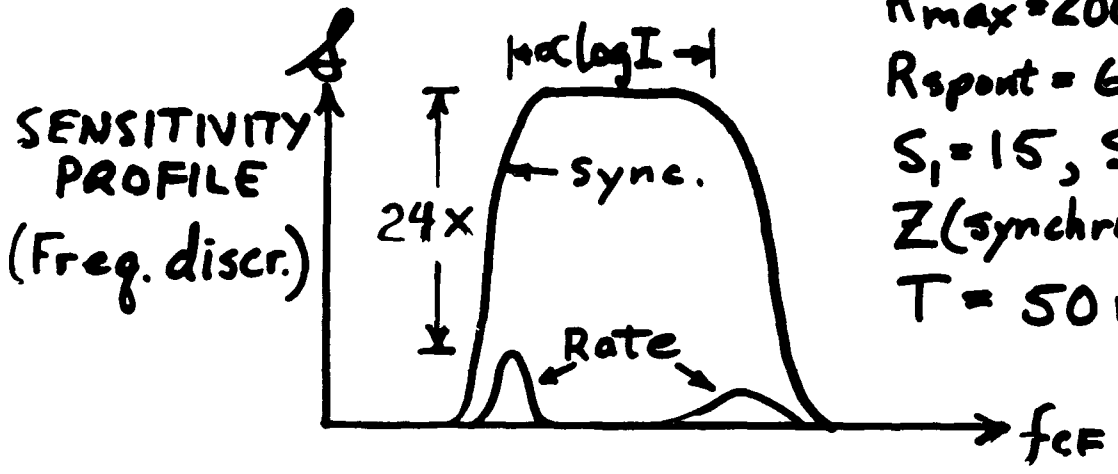
$R_{max} = 200$ Spikes/sec.

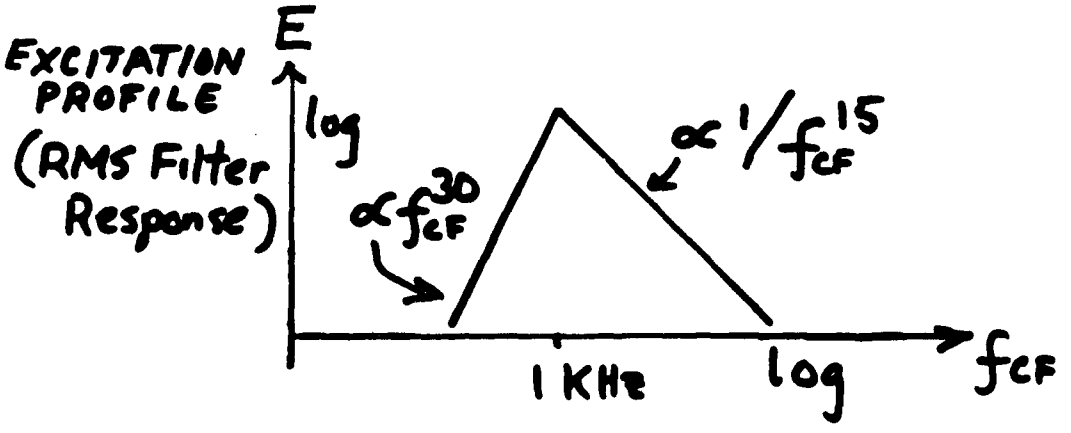
$R_{spont} = 67$ "

$S_1 = 15$, $S_2 = 30$

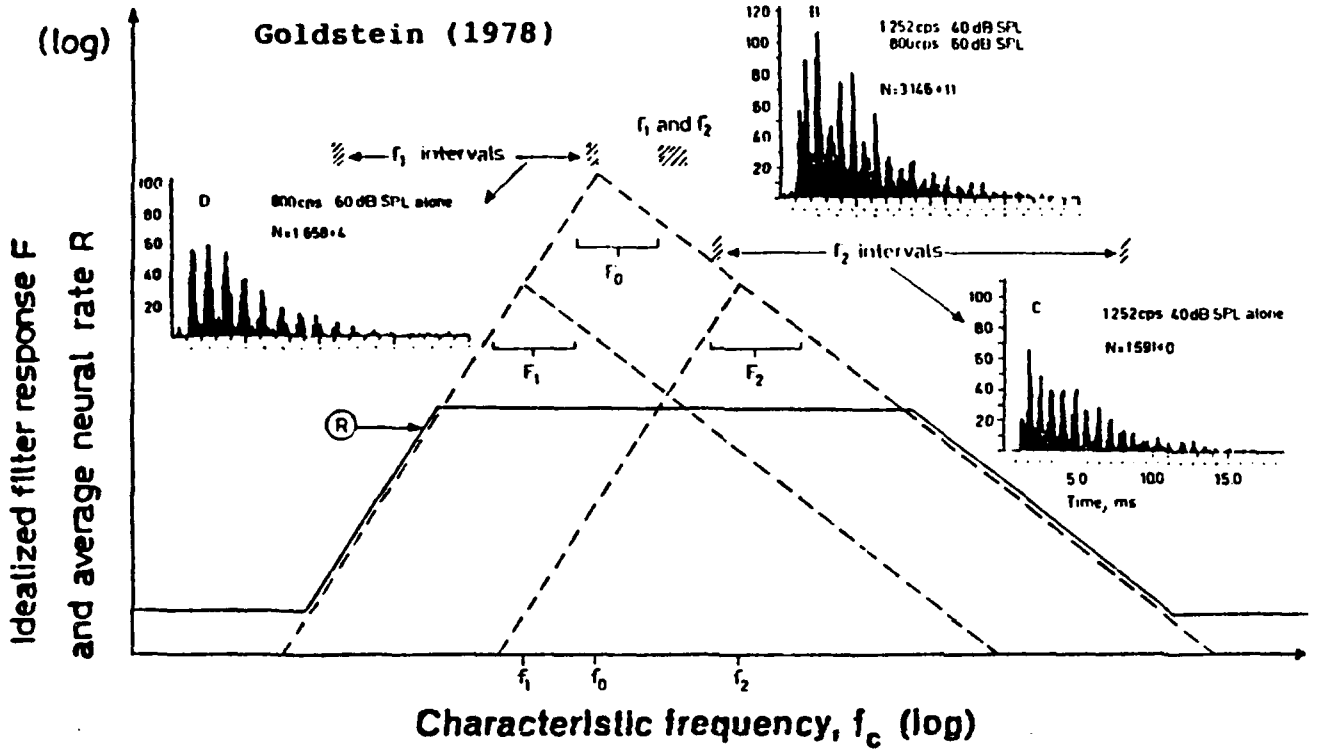
$Z(\text{synchrony}) = 3.2$

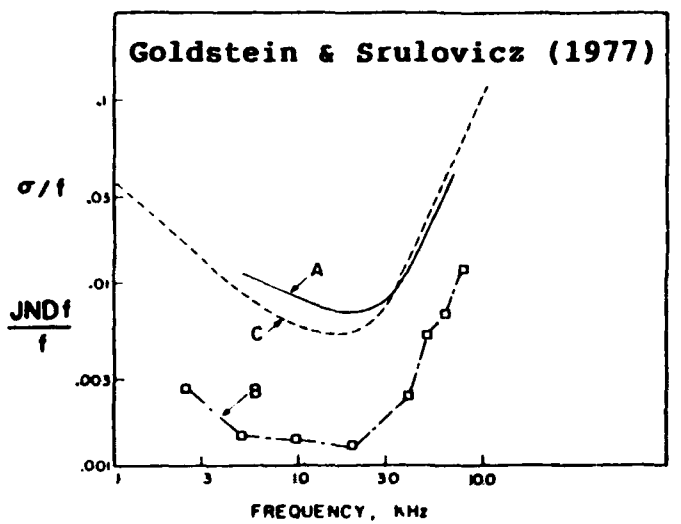
$T = 50$ ms





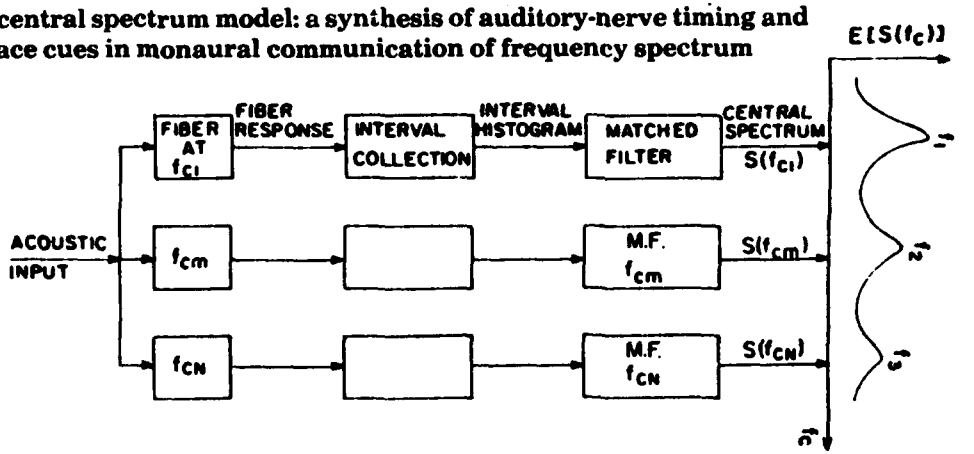
NEURAL REPRESENTATION OF COMPONENT FREQUENCIES





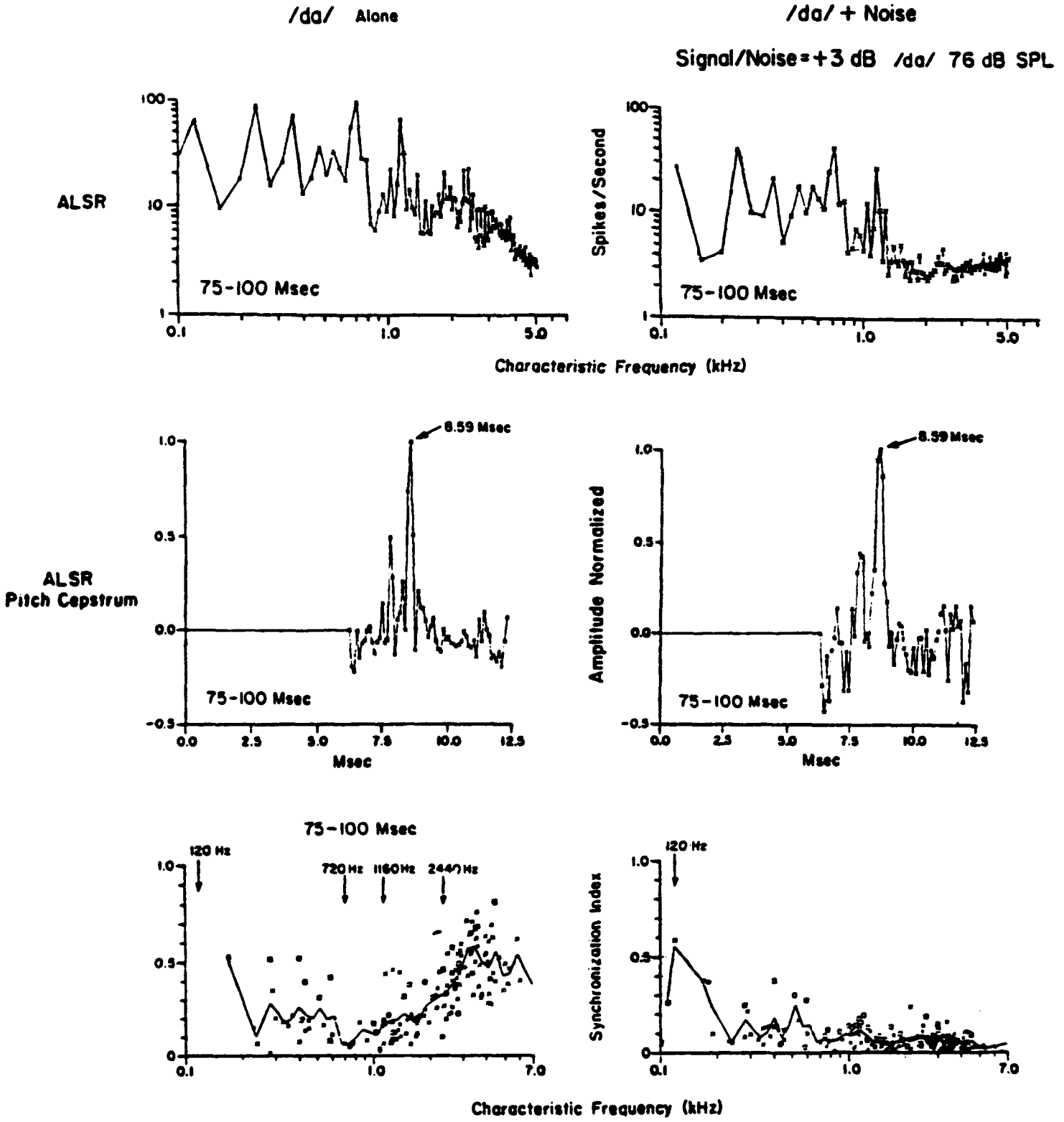
Srulovicz & Goldstein (1983)

A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum

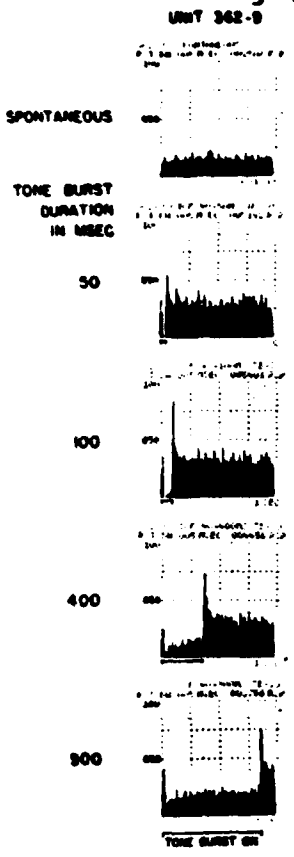


NEURAL REPRESENTATION OF VOICE PITCH

Miller & Sachs (1984)

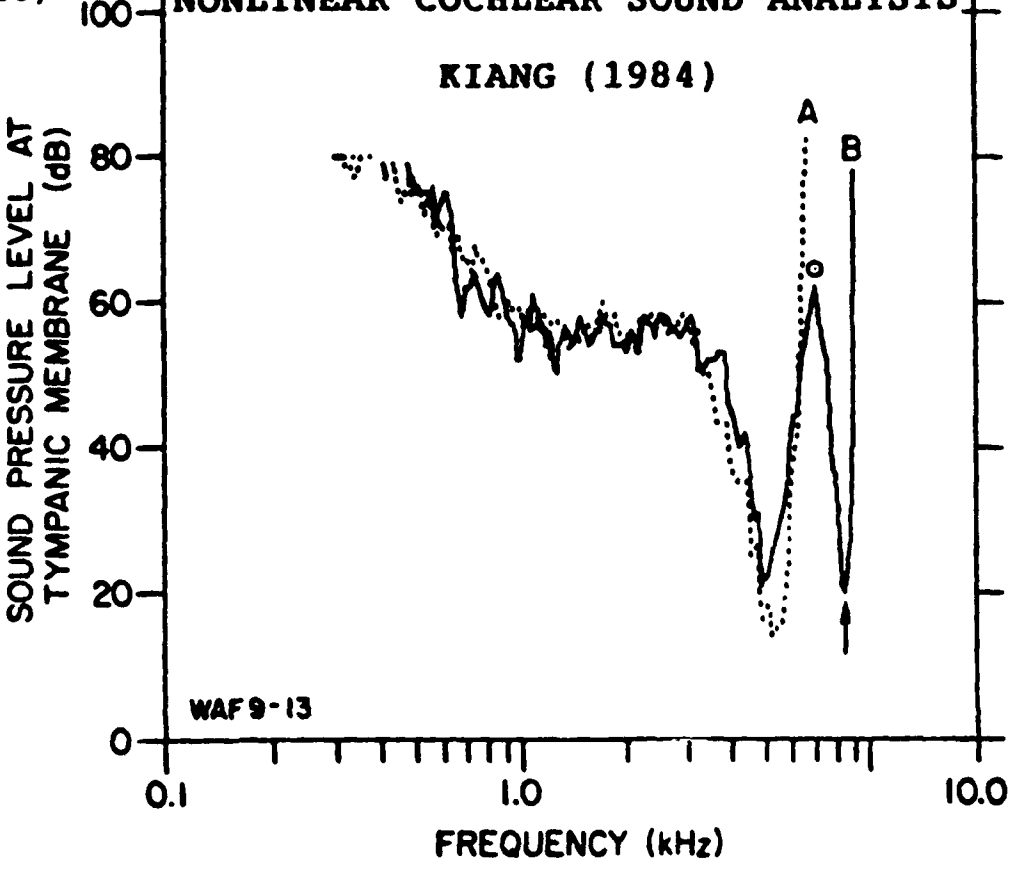


Sachs & Kiang (1968)

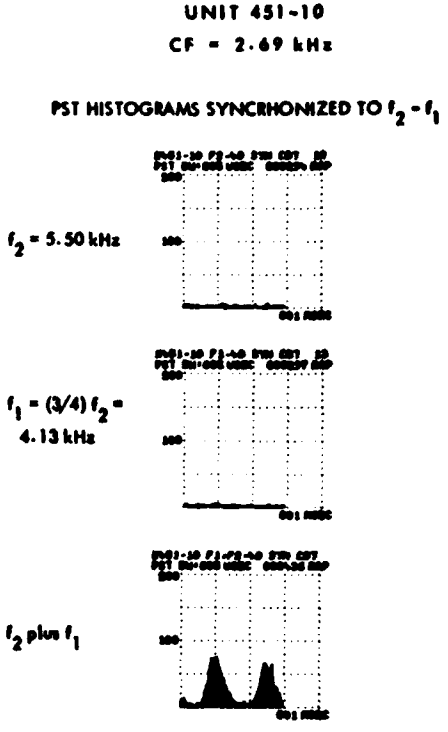


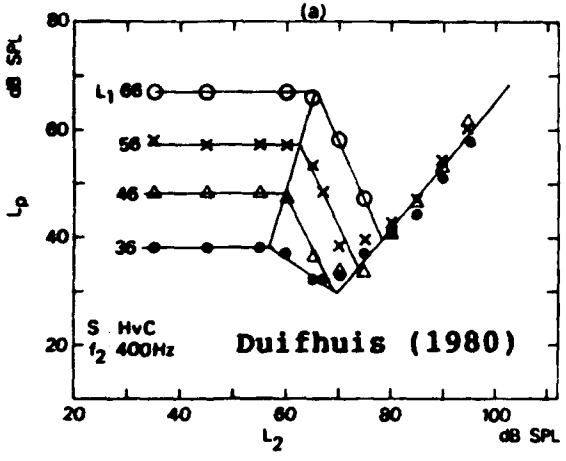
NONLINEAR COCHLEAR SOUND ANALYSIS

KIANG (1984)



Goldstein & Kiang (1968)





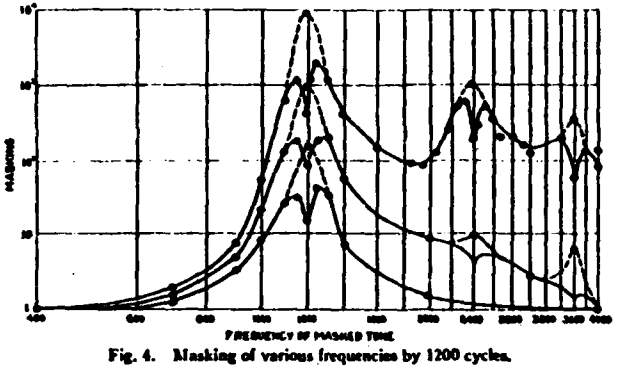


Fig. 4. Masking of various frequencies by 1200 cycles.

Wegel & Lane (1924)

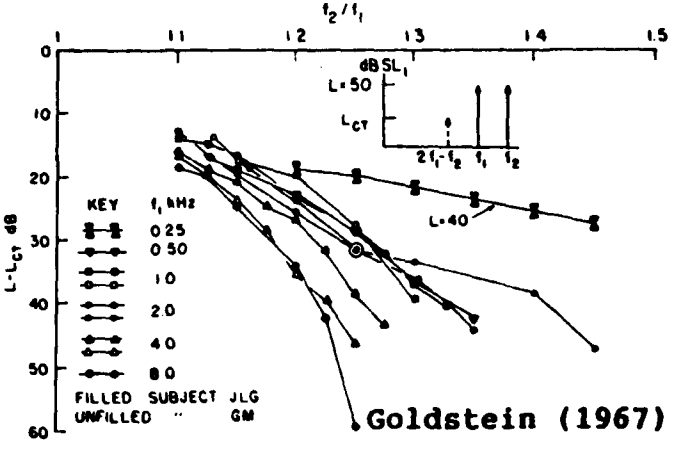
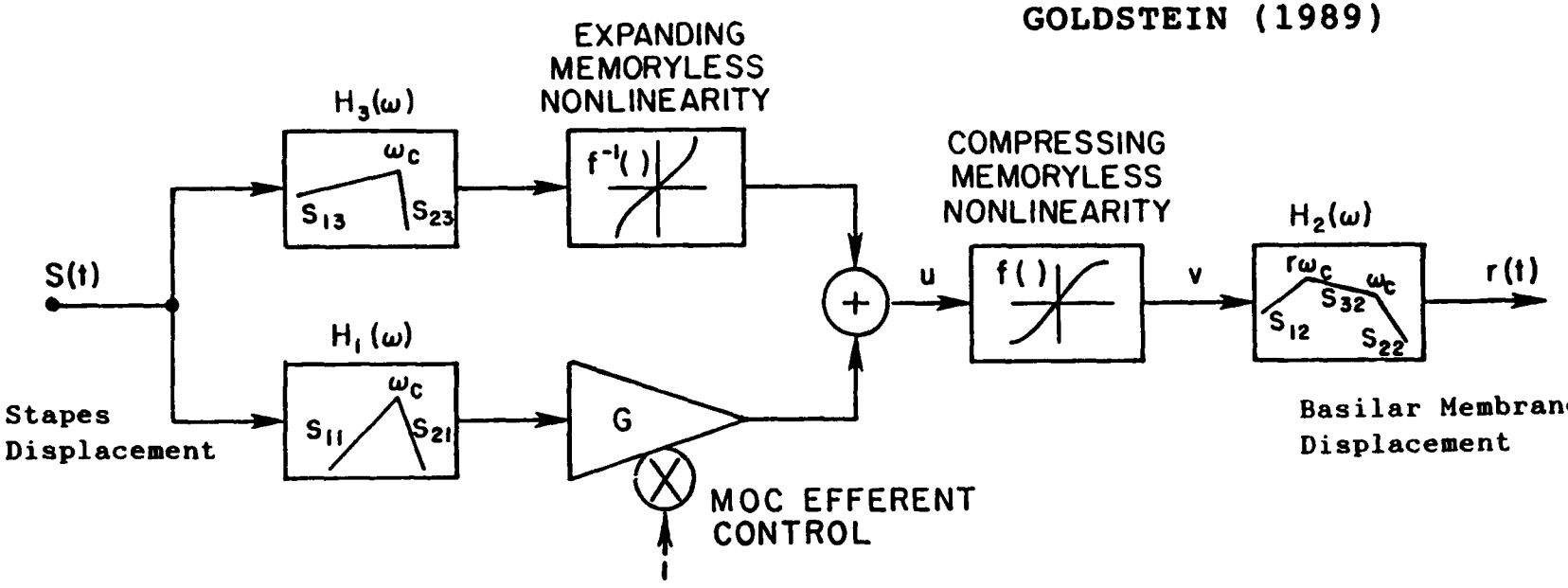
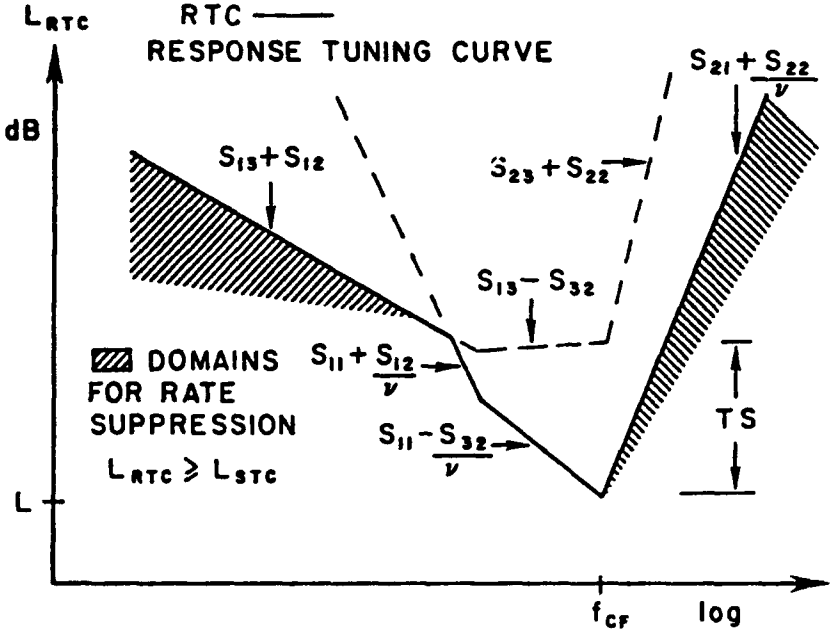


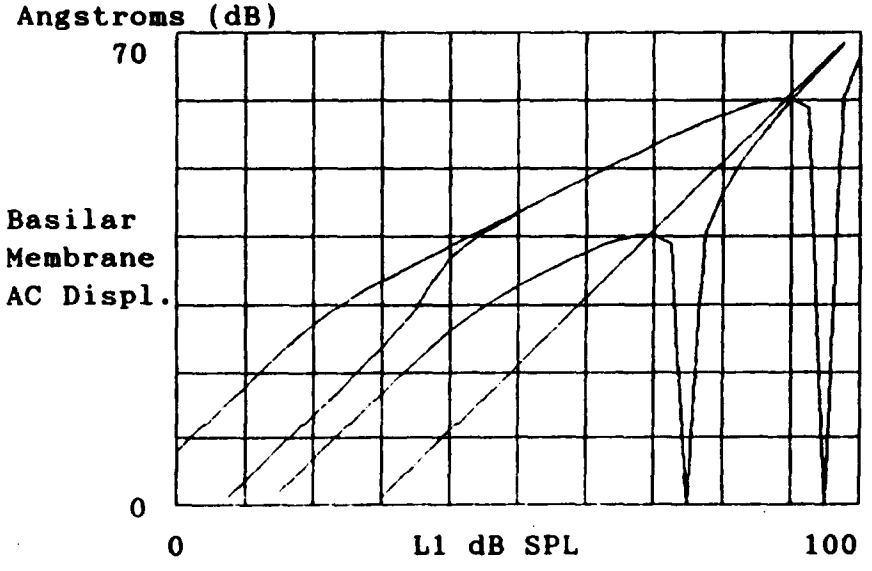
Fig. 7. Level of pitch cancellation tone *vs.* stimulus frequencies.

MULTIPLE BANDPASS NONLINEARITY (MBPNL) FILTER MODEL

GOLDSTEIN (1989)

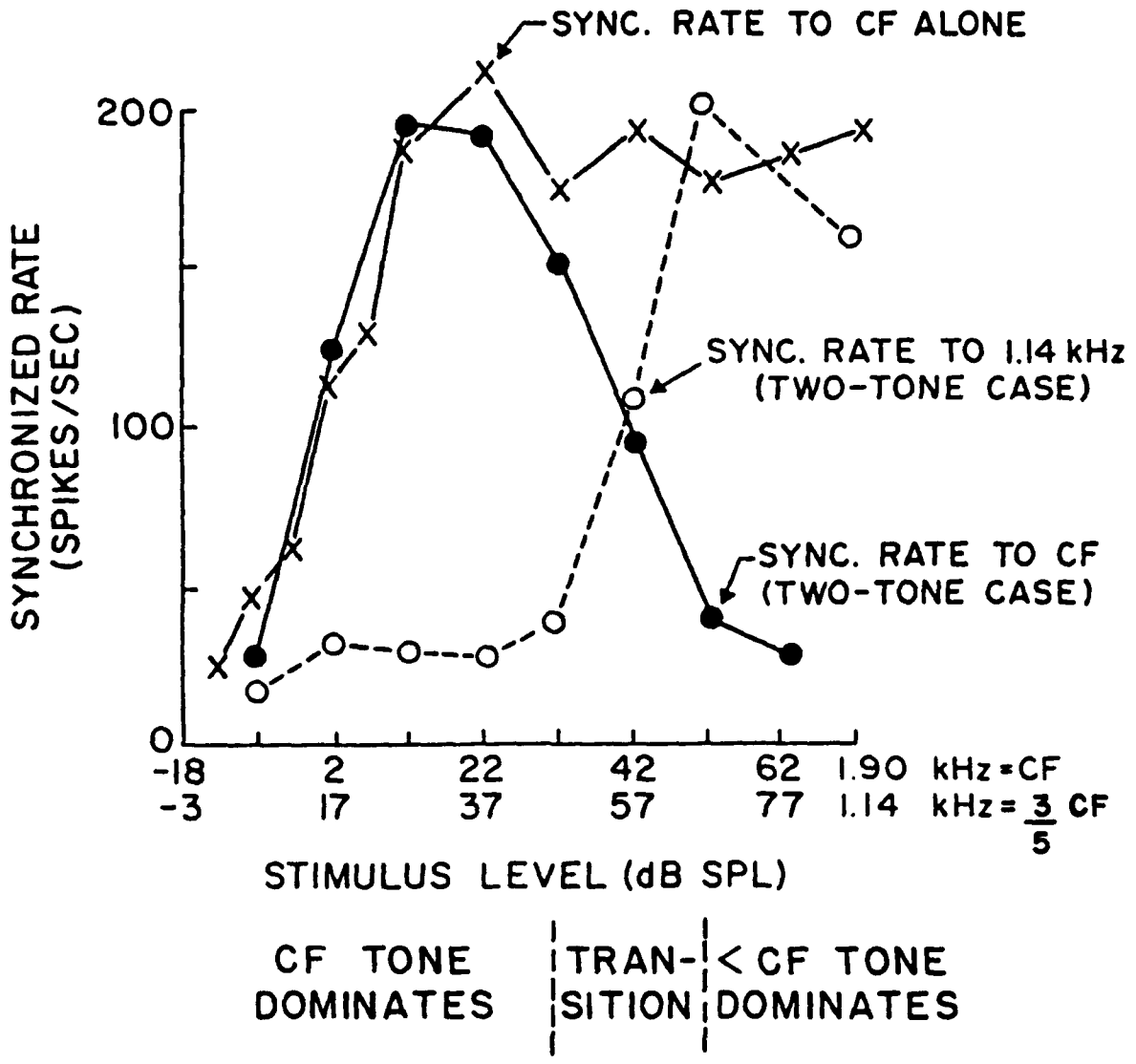






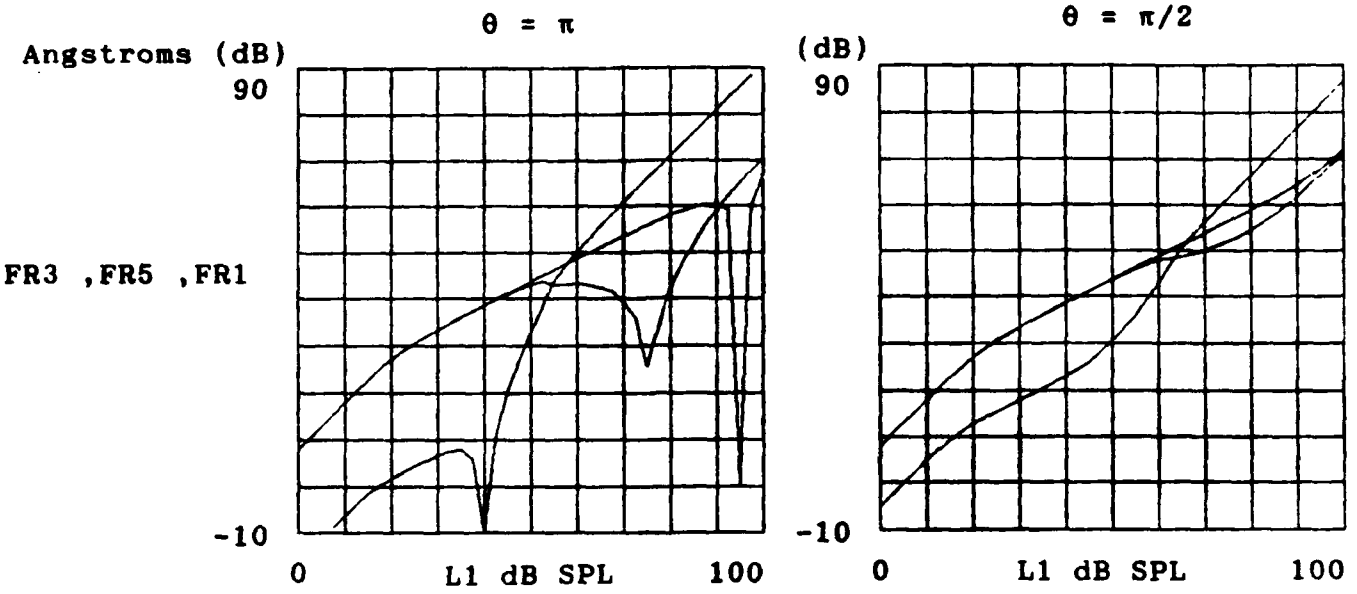
SOUND LEVEL DEPENDENCE OF COCHLEAR FREQUENCY ANALYSIS
"SYNCHRONY SUPPRESSION"
Sachs, Young, Schalk & Bernardin (1980)

FOURIER COEFFICIENTS vs. STIM. LEVEL FOR PERIOD HISTOGRAM RESPONSE TO TWO-TONE STIMULUS

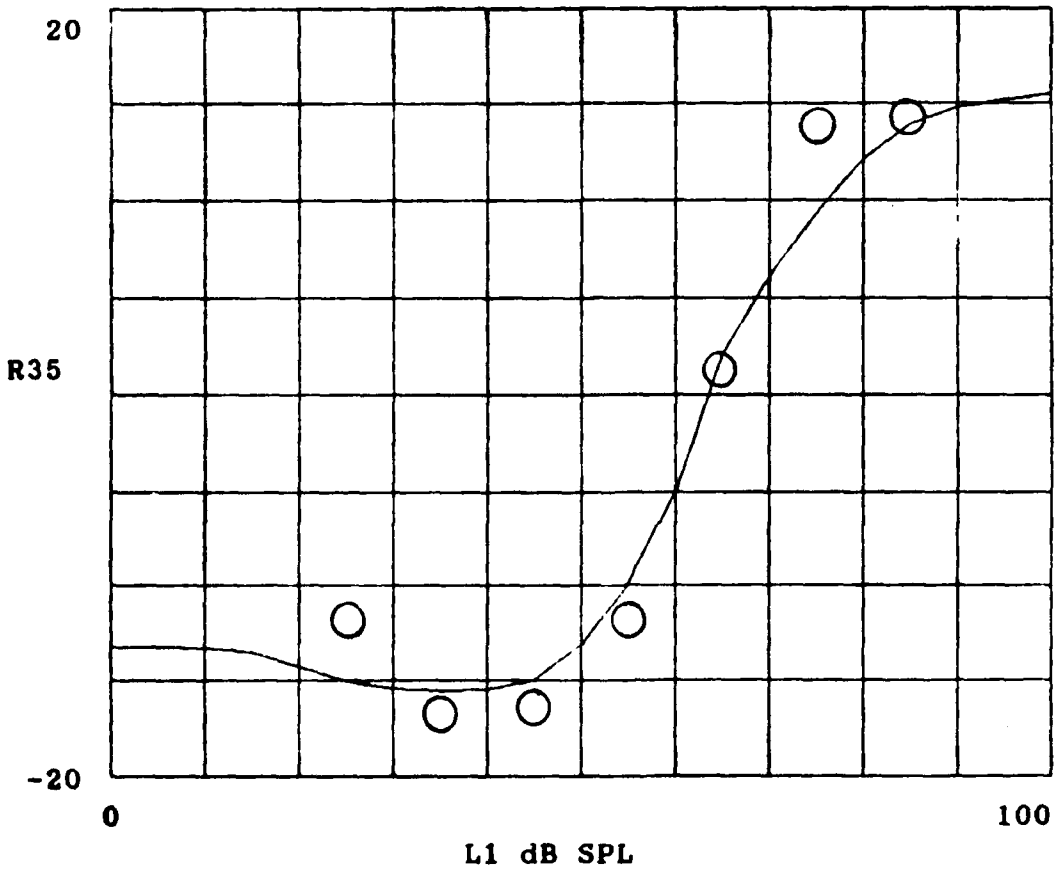


slide 14

MBPNL SIMULATION OF SACHS, et al (1980)

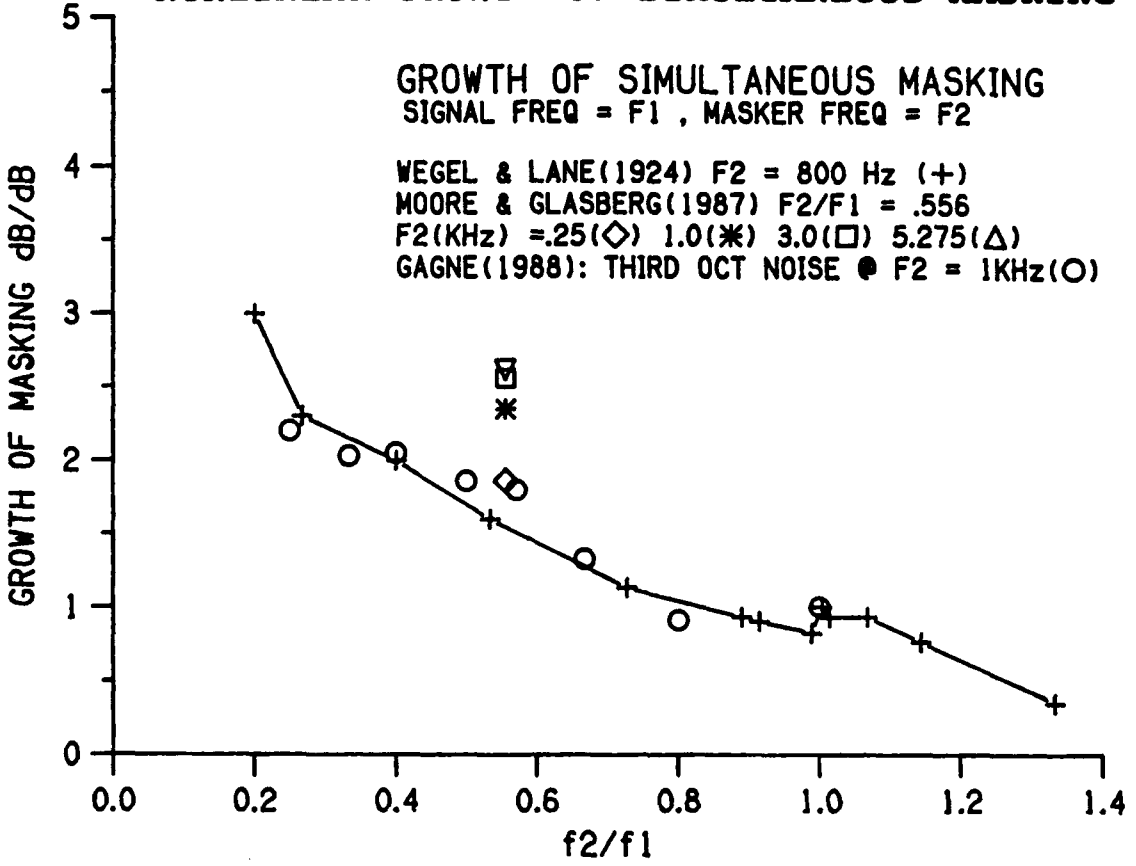


dB Ratio of 3rd to 5th Harmonic Responses: Theory & Data

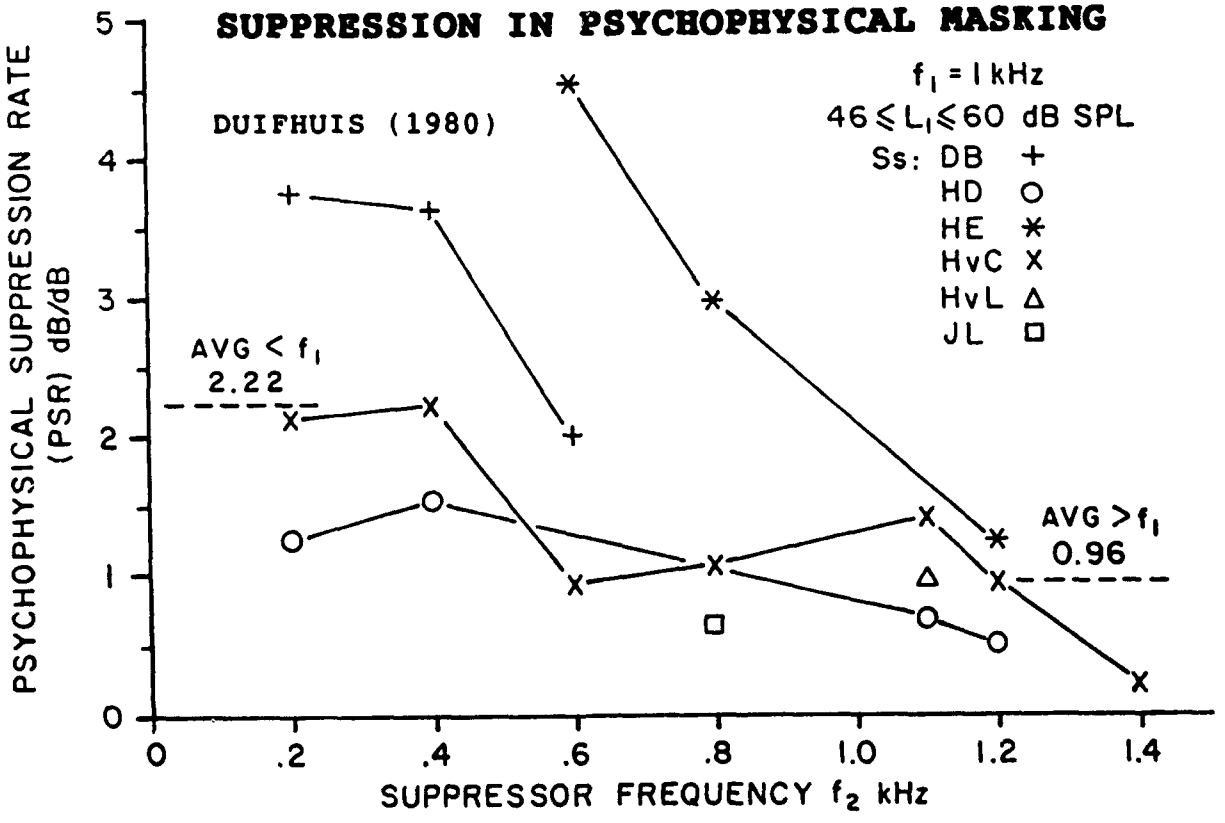


MBPNL Simulation of Sachs, et al (1980) Synchrony Suppression.

NONLINEAR GROWTH OF SIMULTANEOUS MASKING



SUPPRESSION IN PSYCHOPHYSICAL MASKING



J. F. SCHOUTEN

SCHOUTEN'S (1970) RESEARCH TRIANGLE

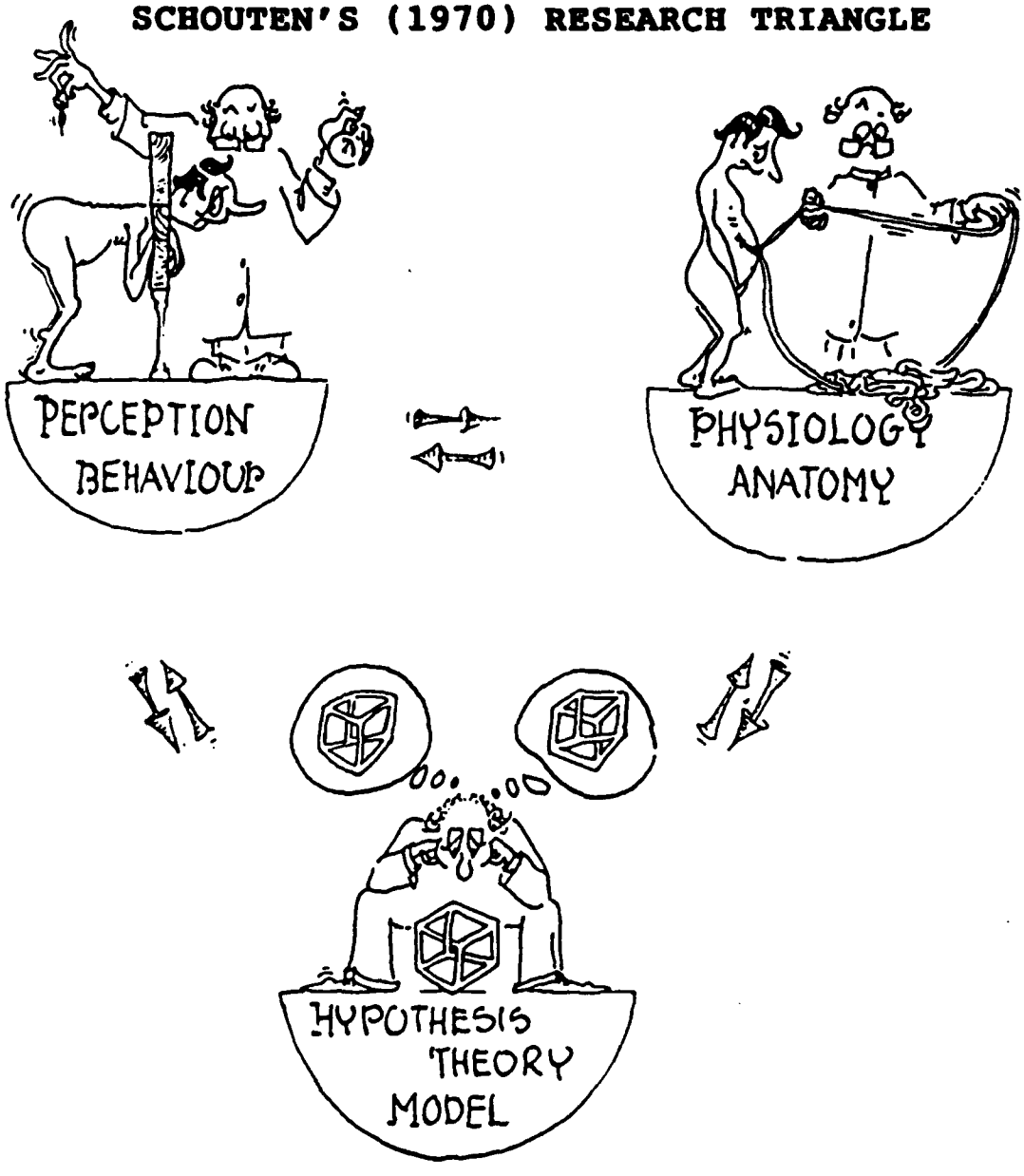


Fig. 1. The eternal triangle in the research on human perception.
Top left corner: Experimental evidence from perception and behaviour.
Top right corner: Experimental evidence from anatomy and physiology.
Bottom corner: Modelmaking, hypothesis and theory.
Along the sides: Mutual interaction.

-- ACCELERATED INTERACTION

Effects of duration on intensity discrimination: Psychophysical data and predictions from single cell response. R.R. Fay, W.P. Shofner and R.H. Dye (Parmly Hearing Institute, Loyola University of Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

Invited paper presented by R. Fay at the special session "Interactions between Neurophysiology and Psychoacoustics," Acoustical Society of America, May 22-26, 1989, Syracuse, New York.

Abstract

An ROC analysis was performed on responses of single auditory nerve fibers (goldfish) and cochlear nucleus cells (gerbil) in order to predict intensity discrimination (in the goldfish and human) as a function of signal duration. To evaluate that the mean and variability of spike counts within single units account for psychophysical performance, spike number distributions were obtained (N=100) for several durations (20 to 400ms) and level differences (0.5 to 4dB) at a unit's best frequency. The percent correct performance based on spike counts was found by generating ROC curves from empirical distributions and computing the area under the ROC (P(A)). Theoretical psychometric functions were compared with psychometric functions from human and goldfish listeners obtained using a 2IFC paradigm (human) and a rating method in classical respiratory conditioning (goldfish). The forms of the neural and psychophysical duration functions are similar in the mammal and the fish, but the fish shows higher thresholds compared with the human and with the neurophysiological predictions. In general, psychophysical performance is well modeled by the optimum processing of spike counts from individual cells. [Supported by a Center Grant from NINCDS]

I am going to speak to you today about an approach we have been taking toward the further understanding of neural codes or representations which underlie sensory behavior and perception. This is the general goal of our work - to understand the dimensions of neural activity that are actually used by the brain in making decisions and in perception.

Specifically, I will be speaking about data from psychophysical studies on intensity discrimination in fishes and the human, and on neurophysiological studies of the response properties of single auditory nerve fibers in the goldfish, and of cochlear nucleus cells of the gerbil. The primary question concerns the evidence that the organism uses in judging differences and changes in sound level. From this point on, I will use the term "intensity" to refer to sound level, even though the use of this word is technically incorrect to the physicist or acoustician.

The question of sound intensity perception is perhaps the most basic question in investigations of sensory behavior. In general, we know that the magnitude of the perception grows as the stimulus intensity grows, and the precise form of this relationship has been investigated by psychologists since the last century. Closely related is one of Adrian's fundamental observations of sensory physiology - that the neural response of peripheral sensory fibers grows as stimulus intensity grows. This gives us a simple, qualitative correlation between the magnitude of the neural response and the magnitude of perception.

I am interested in looking closely at this relationship between neural activity and behavior. In what sense does the growth of the neural response <u>cause</u> a change in perceived sound intensity? -- What is the neural code for sound intensity? What evidence does the brain use in judging differences and changes in sound intensity? In this paper I focus exclusively on the notion that activity within single neural channels underlies intensity discrimination. Othe alternatives, for example that the number of active cells, or the activation of cells with different thresholds, or the sum of all neural activity underlies intensity discrimination are not considered here.

To approach this kind of question, we need behavioral measures of sound intensity perception and physiological measures of the neural response determined for the same species under the same stimulus conditions. In addition, we need an hypothesis about what dimensions of neural activity likely underlie sound intensity perception, and an hypothesis about how decisions are made on the basis of this information. Finally, we need experimental paradigms which allow the results of these two kinds of experiments to map on to one another.

I'll begin with the behavior. Of the many aspects of sound intensity processing that have been studied, we have concentrated on discrimination acuity. The question is - what is the smallest change in sound intensity that can be detected? This question asks about the limits of system performance, and essentially focuses on the causes of discrimination error. We measure the limits of performance and then ask what physiological factors determine these limits.

<u>PSYCHOPHYSICS</u>

Psychophysical studies of sound intensity discrimination in the goldfish (*Carassius auratus*) make use of the classical conditioning of respiration. A head-tail electric shock produces an unconditioned respiratory suppression that

is measured with a thermistor. Any stimulus which precedes the shock by several seconds eventually takes on some of the respiration suppression effects of the shock, and the conditioned suppression becomes an indication that the fish perceives the stimulus. Fig. 1 illustrates that this method has been used in a variety of experiments to study sound detection, intensity and frequency discrimination, masking, temporal discrimination, etc. Focus here on the intensity discrimination case - the fish hears a continuous series of identical brief tones. The signal is a change in the intensity of the tones lasting for several seconds. If the change in intensity is detected, respiration is suppressed. Fig. 2 shows how we quantify the degree of suppression as a ratio of respiratory activity before and during the signal. In the past, we have used an automated threshold tracking method to estimate the intensity difference which is just detected - and we end up with the intensity discrimination threshold (IDT), in decibels. This is the smallest decibel change in sound intensity required for a reliable respiration suppression response.

In our previous studies on the goldfish, we have looked at the effect of sound duration on intensity discrimination (Fay, 1985). Without going into the details of these different functions, Fig. 3 shows that between 20 and 500 ms, the IDT declines. The longer the duration of the sound, the greater is the acuity for intensity discrimination. These are power functions (log IDT in dB vs log duration) with slopes of about -0.33.

The IDTs of human listeners were also measured at 250, 1000, and 4000 Hz, at 70 dB SPL, for signals of durations of 20, 50, 100, 200, and 400 ms using a two-interval, forced-choice paradigm. In both the goldfish and human studies, rise/fall times were 10 ms. Fig. 4 shows the results from one subject (RHD), including psychometric functions and the 76% correct thresholds derived from them plotted as a function of duration. Although the well-trained human listener is more sensitive to intensity change than the goldfish, the duration functions are similar, showing power functions with approximately the same slope (See also Florentine, 1986).

Figure 5 shows the goldfish and human data together with IDTs determined in a number of experiments reviewed by Fay (1988) on intensity discrimination in several fish, bird, and mammal listeners (unconnected points). The points have also been projected to the right to give the impression of the distribution of IDTs obtained in animal psychophysical studies.

In investigating the neural causes of these behaviors, we have initially focused on the size of the threshold at one frequency (200 Hz) for the goldfish, and at the best frequency of cochlear nucleus cells for the gerbil (*Meriones unguiculatus*). Of primary interest is the effect of signal duration on the IDT.

NEUROPHYSIOLOGY

Physiological observations on this question begin with the simple relation between sound intensity and the number of spikes evoked within single cells of the auditory system.

In the goldfish, the saccule is the primary auditory receptor organ, and we have recorded from fibers of the saccular nerve that respond best at 200 Hz (Fig. 6). These functions show the number of spikes evoked during a 100ms tone presentation as a function of the sound pressure level for several auditory nerve (saccular)

fibers of the goldfish. Each point plotted is the spike count for a single tone stimulus presented once at intervals of 1dB.

There are two obvious and important points illustrated here. First, increasing sound intensity causes an increase in the spike count within the dynamic range of a given fiber. The rate of growth (slope) of spike count varies from one fiber to another. Second, the spike response is variable, giving rise to the jagged form of some of these curves. Some fibers are more variable than others.

Here is illustrated the essential problem facing the nervous system in making decisions about sound intensity based on the spike response of individual nerve fibers. Imagine that you are a homunculus within the brain and that you are presented with the spike count from a single fiber for a single tone presentation. As in Fig. 7, you have to decide whether the sound intensity of the tone is low (left column), or high (right column). All the evidence about sound intensity you have is the spike count received from one fiber. What would be your best possible performance? A simple and rational strategy to take would be to adopt some criterion spike count - and to say "HI" if the received count exceeded the criterion, and to say "LO" if the count fell below the criterion. In doing this, the percentage of times you would be correct or incorrect in this judgement depends on: 1) the intensity difference between the low and high intensity tones, 2) The slope of the spike rate - intensity function, and 3) The variability of the fiber in generating spikes.

The Theory of Signal Detectability (TSD) gives us a simple way to relate these variables and predict the performance of an ideal decision-maker in this task. To do this, we start with two distributions of spike counts - one for the low intensity sound, and one for the high intensity sound (Fig. 8). Placing a decision criterion somewhere along the spike count axis allows us to compute the proportion of times we would be correct in detecting the high intensity sound (shaded area of the distribution to the right - HITS), and the proportion of times we would say "high" when, in fact, the low intensity sound was presented (shaded area of the distribution to the left - FALSE ALARMS).

For every possible decision criterion, the two values - HIT RATE and FALSE ALARM RATE - can be computed and plotted against each other. Connecting these points with a line defines the so-called Receiver Operating Characteristic (ROC). Notice that if the spike count distributions for the low and high intensity tones overlap completely, the HIT RATE and the FALSE ALARM RATE will always be equal, and the ROC will fall on the diagonal line. If the two distributions don't overlap at all, the ROC will be two straight lines coinciding with the left and top sides of the ROC space. Any intermediate overlapping of the two distributions will result in a ROC intermediate between these two extremes.

Now, TSD tells us that the best prediction of the performance of an ideal observer deciding about sound intensity on this basis is made by integrating the area under the ROC, [P(A)]. The area under the diagonal line is 0.5, indicating that one would be 50% correct in deciding about sound intensity in a 2IFC paradigm. This is equivalent to guessing in a situation containing no information. The area within the entire ROC space is 1.0, indicating that one would be 100% correct if the distributions did not overlap at all.

An area of 0.76 under the ROC can be taken as "threshold" performance. The intensity difference between the low and high intensity tones leading to 76%

correct performance can be defined as the predicted optimum IDT for this system. Remember, the "system" in this case consists of a single fiber and an optimum decision process using spike count as the decision variable.

Now, we can apply these ideas to neurophysiological data simply by obtaining the spike count distributions for two different sound intensities. The forms of these distributions (their means and variabilities) then allow us to predict the performance of an ideal decision-maker in discriminating between the two sound intensities. This prediction can be viewed in two ways - one conservative and one liberal. The conservative view says that this analysis is useful primarily in defining how "significant" any particular change in mean spike count is, or how well any given fiber represents a given sound intensity difference in terms of spike count. The liberal or optimistic view is that the predicted performance from a single fiber could be used to estimate the performance of the whole system (the organism) in sound intensity discrimination. The initial assumptions here are: 1) that decisions can be made on the basis of the spike counts within a single fiber; 2) that the organism operates like an ideal observer; 3) that spike count is the decision variable.

Now I want to show you some neurophysiological data so that we can begin to decide whether we might want to entertain the optimistic view. The experiment begins by obtaining a spike rate -sound intensity function (Fig. 9). Some intensity is selected, and we present 200, 200 or 400 ms, tone bursts once per second. 100 of the tones are at the intensity selected (the "low" intensity), and 100 are at a slightly higher intensity (the "high" intensity). The spikes evoked by each stimulus are counted 4 ways - we get the count over the first 20ms, 50ms, 100ms, and 200ms (and 400 ms for the gerbil). Spike count distributions are formed for the high and low intensity sounds at each of the 4 or 5 measurement durations. Each high low intensity pair of distributions are analyzed to give us the ROC function, and the areas under the ROCs are computed. These values are the predicted performance of the ideal observer in discriminating between the high and low intensities for the 4 different listening durations.

The series of 200 bursts are repeated with the same Low intensity values but with different high intensity values. With these data we obtain the functional relation between high-low sound intensity difference, and predicted performance for each listening duration. Sample neural "psychometric functions" are shown in Fig. 10. Clearly, performance improves as the high-low intensity difference increases, and as the listening duration increases. Each of these functions defines a predicted IDT for this nerve fiber (the intensity difference corresponding to 76% P(A)).

Figure 11 shows a comparison between two cells (one from the goldfish auditory nerve and one from the gerbil cochlear nucleus) and the performance of two human listeners (one well trained, and another untrained). The two cells were selected because they showed good IDT sensitivity. The main point here is simply that the duration effect is similar for the two cells and for the human listeners, and the general form of the functions are qualitatively similar.

I would now like to say something briefly about the diversity of responses from different cells of the gerbil cochlear nucleus. Fig. 12 illustrates the performance of 4 cochlear nucleus cells identified as transient choppers (chop-t). These cells were selected to show that this cell type produces quite

different patterns of performance in representing intensity differences. Some cells show "flat" duration functions, some show non-monotonic duration functions, and others (the majority) show monotonically decreasing ITDs with duration. Thus, patterns of intensity discrimination performance can vary widely within a cell type.

Fig. 13 shows performance for 4 different types of cells identified as transient choppers, primary-like, on-L types, and transient type IV. In spite of these differences in classification (based on the shapes of PST histograms), these various cells can show intensity representation performance that are remarkably similar. One generalization that can be made is that sustained choppers (not shown) are generally the most sensitive to sound intensity differences.

Figure 14 shows the duration functions for about 25 gerbil cochlear nucleus cells of various types, stimulated at an intensity giving the best intensity representation performance. The shaded area represents the middle 85% of the goldfish auditory nerve cells for comparison. In general, some cells are quite "flat," while others are not. Clearly, some cells represent intensity differences quite well, and others do not.

Figure 15 is a similar plot for the goldfish auditory nerve cells, with the shaded area representing the middle 85% of the gerbil cochlear nucleus data. In general, we can say that there are no "flat" cells in the goldfish auditory nerve, and that the goldfish functions are steeper than those for the gerbil at short durations. On the other hand, both sets of cells (Figs. 14 and 15) show a similar pattern of duration dependence.

Figure 16 compares the goldfish auditory nerve and the gerbil cochlear nucleus cells in terms of central tendency (the middle 85% and their medians). There is similar sensitivity at long durations, but the goldfish show reduced sensitivity at the short durations. It is not yet clear whether this difference is characteristic of the species, the frequencies used (200 Hz for the fish, kHz range for the gerbil cells), or of the level in the nervous system at which measurements were made (auditory nerve versus the cochlear nucleus).

Figure 17 compares, from earlier figures, the psychophysical results with the neural results. In general, the human psychophysical data coincide quite well with the most sensitive of the gerbil cochlear nucleus cells. On the other hand, the goldfish psychophysical data are generally poorer than the best of the goldfish auditory nerve fibers. Notice, too, that most mammals (and birds and fish) perform less well than the best gerbil cochlear nucleus cells.

This pattern of results might lead one to the conclusion that psychophysical performance in the human is based on the small number of cells which represent intensity changes best (assuming that gerbil physiology is like human physiology). This might also suggest that the superior performance of the well-trained human listener is based on the ability to <u>select</u> channels with superior information, or to <u>combine</u> information across channels in ways that non-human listeners cannot.

Alternative explanations for the species differences in the relationships between the psychophysics and neurophysiology are: 1) Spike count, as we have measured it, is not the neural code, or the "real" decision variable used in making intensity judgments; 2) Spike count may be only one of the codes used; 3) We

may not have designed the animal psychophysical paradigm properly to estimate optimal performance (e.g. the amount of uncertainty in the tasks devised for humans and animals may not be the same); or 4) Psychophysical performance may be more closely related to some ensemble of neural representations rather than to the performance of individual nerve cells.

Taken at face value, the present data show that the intensity discrimination performance of non-human animals is poorer than the "best" cells would predict. This suggests that the information contained within the neural population is not combined in an optimal way by the brain. On the other hand, well-trained human listeners show performance at least as good as the best cochlear nucleus cells of the gerbil, and thus suggest that humans may be able to select channels containing the best evidence. Are the species differences in psychophysical performance due to differing capacities for channel selection, or to capacities for combining information across channels? What is the effect of training human psychophysical observers? Is it to refine capacities for selection or combination, or is it to focus attention on additional cues not considered in this paper?

Figure 18 illustrates that the differences in performance between a well-trained and untrained human listener are greater than the differences between the performance of cells from the auditory nerve of the fish and the cochlear nucleus of the gerbil. The data from the untrained observer (and the non-human animals) may lead us to the view that decisions are not made optimally on the basis of single channel information. The data from the trained observer may lead to the view that optimal decision-making is possible using evidence from selected channels. Which, if any, of these processes actually underlie sensory information processing?

In any case, I think that it will be useful and informative to include animal psychophysical data in the development of hearing models based on neurophysiological data in the future. This may lead to a greater understanding of the differences between humans and non-humans in sensory information processing, and to a better understanding of both human and animal hearing.

References

Fay R (1988) Hearing in Vertebrates: A Psychophysics Databook. Hill-Fay Associates: Winnetka, IL.

Fay R (1985) Sound intensity processing in the goldfish, J Acoust Soc Amer, 78, 1296-1309.

Florentine M (1986) Level discrimination of tones as a function of duration, J Acoust Soc Amer, 79, 792-798.

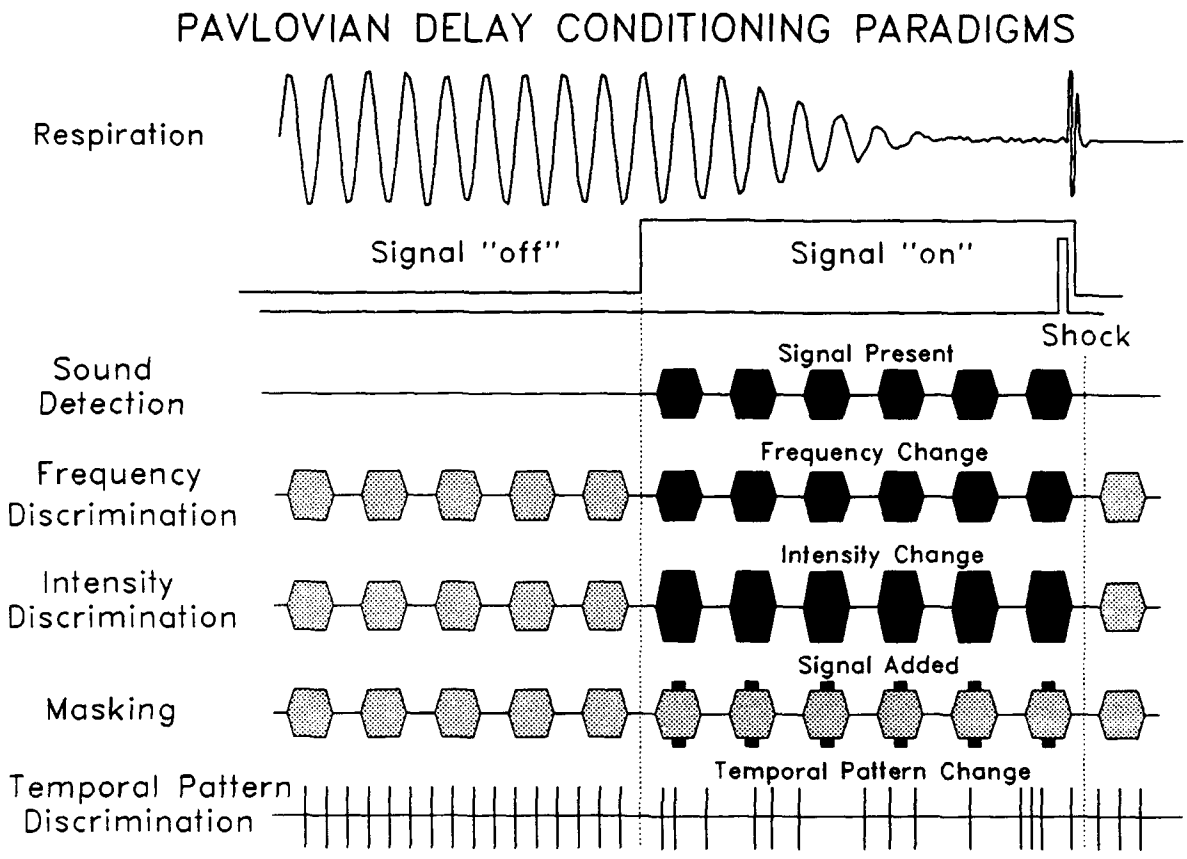


Fig. 1 A classical respiratory conditioning trial for the restrained goldfish illustrating the timing of acoustic signal and shock. The bottom traces illustrate some of the different kinds of detections and discriminations that can be investigated using this method.

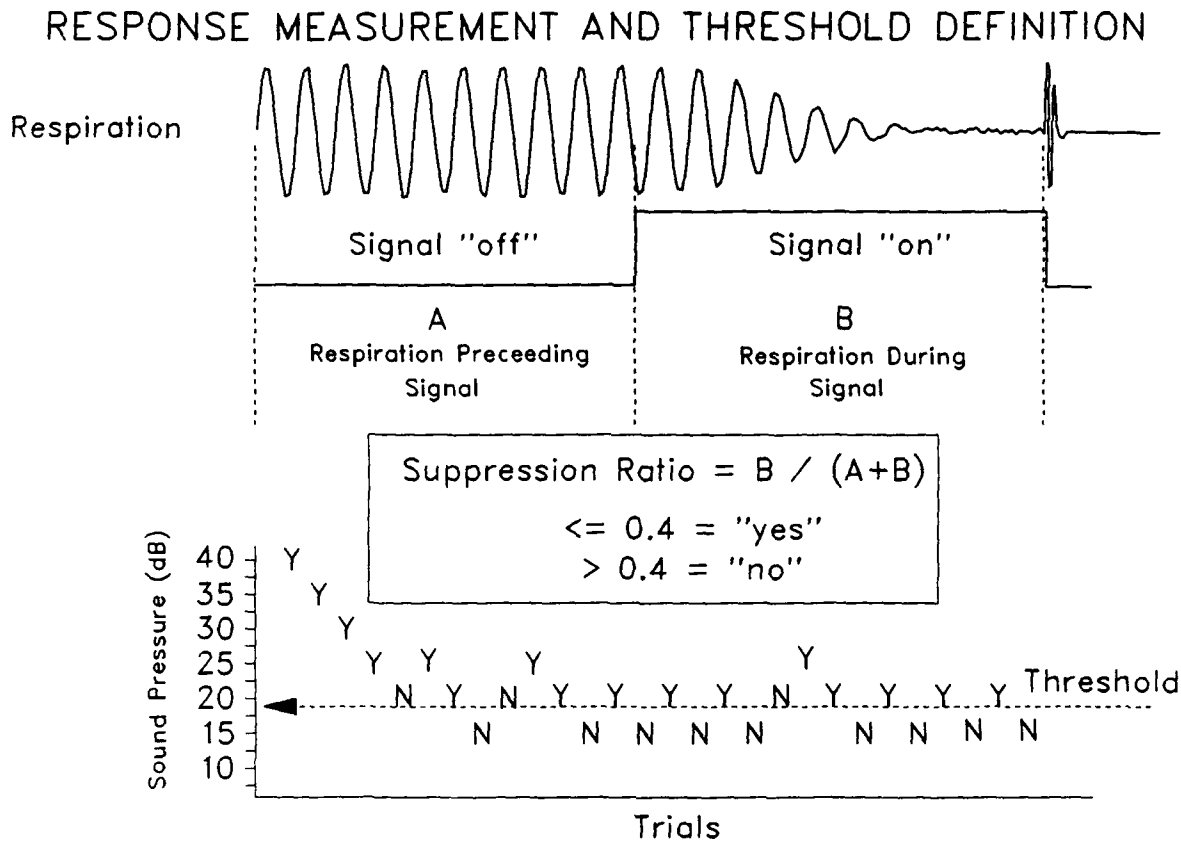


Fig. 2. Illustration of the tracking psychophysical procedure used to obtain intensity discrimination thresholds for the goldfish.

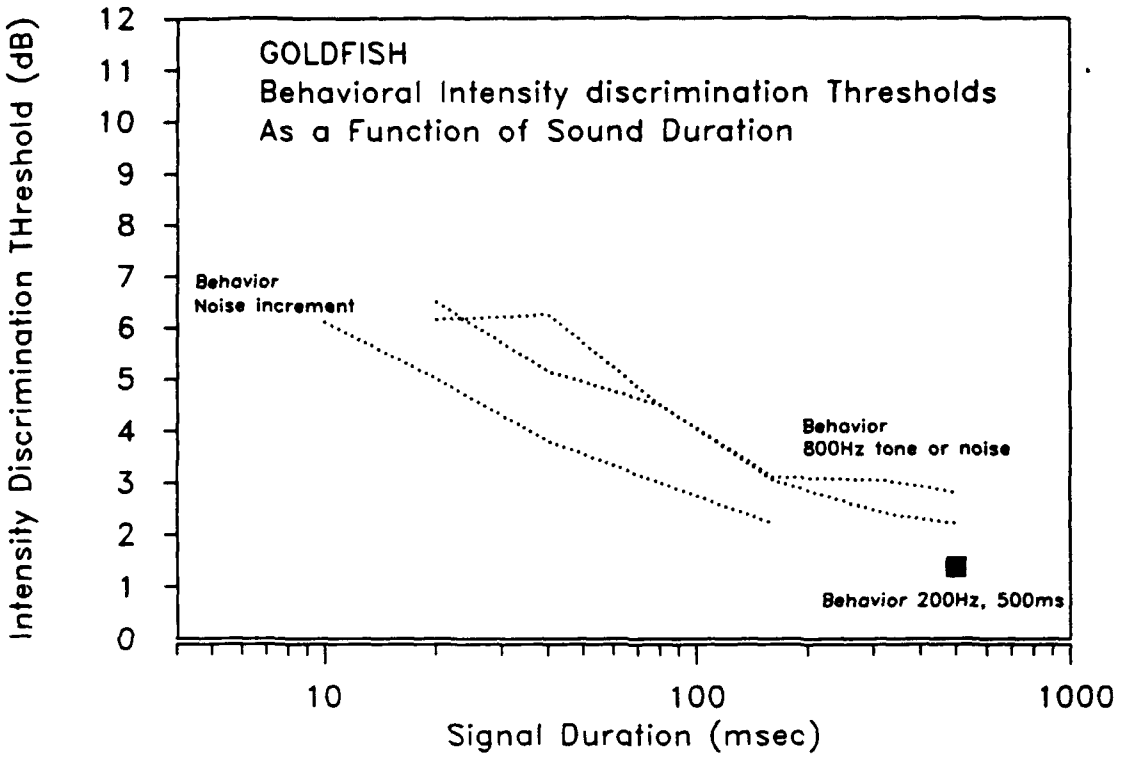


Fig. 3. Intensity discrimination thresholds for the goldfish as a function of signal duration. The dashed lines show the duration effect for increments in continuous noise, for pulsed noise, and for a pulsed 800Hz tone. The filled symbol shows the IDT for a 200 Hz pulsed tone. [Note that all physiological data on the goldfish in this report were obtained using a 200Hz tone. The dashed lines indicate the probable form of the duration functions for 200Hz pulsed tone, but probably give higher thresholds than would be obtained at 200Hz.]

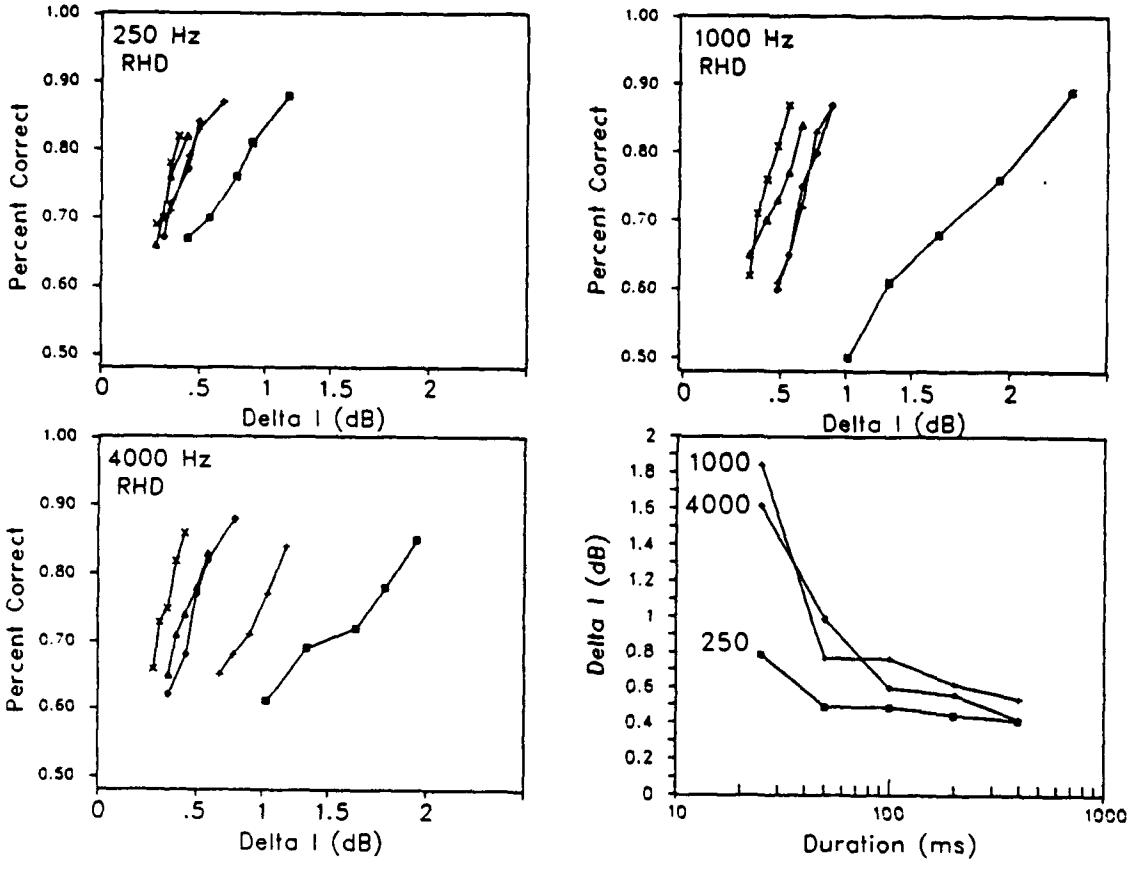


Fig. 4. Psychometric functions and IDT functions of duration for one, well-trained human listener.

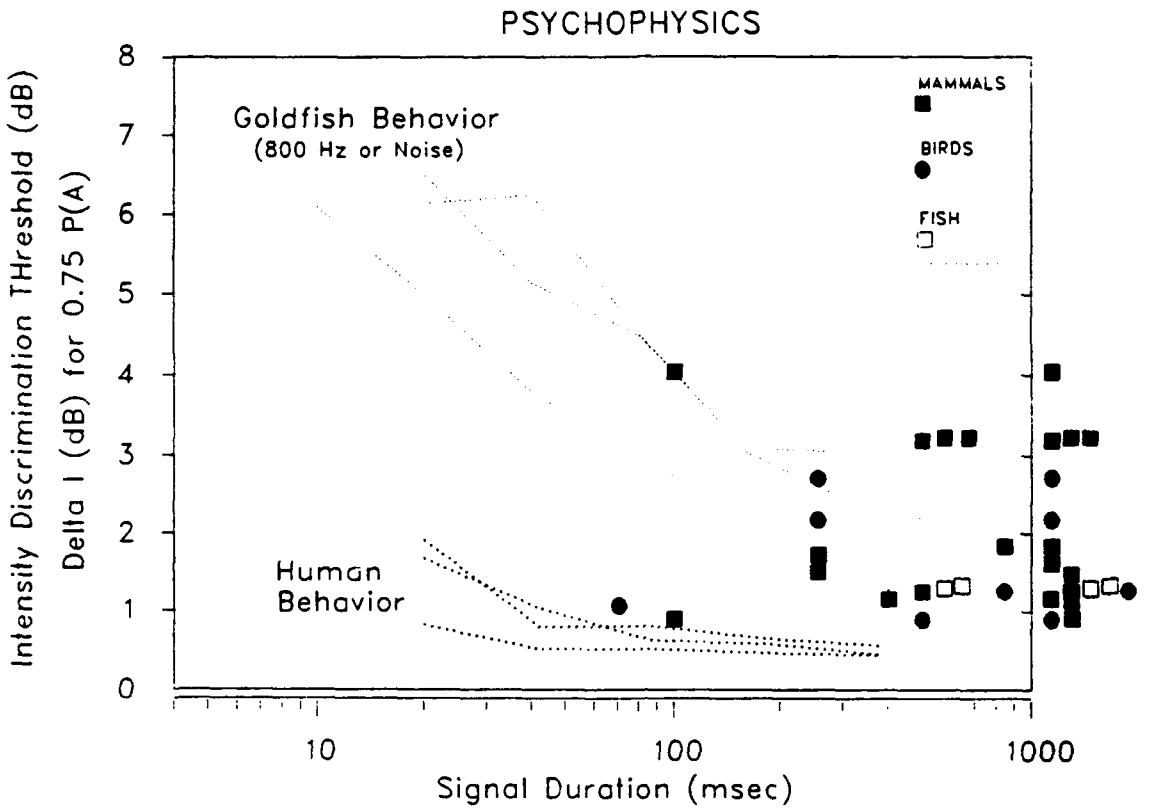


Fig. 5. Comparison of human, goldfish and other animal (see text) intensity discrimination thresholds as a function of signal duration.

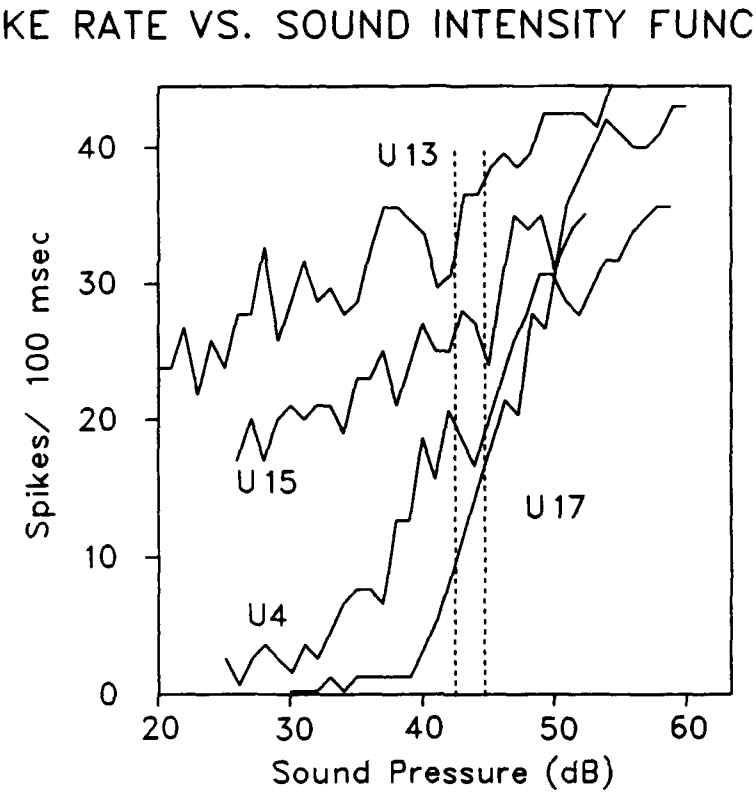


Fig. 6. Spike count - sound intensity functions for several representative auditory nerve fibers from the goldfish.

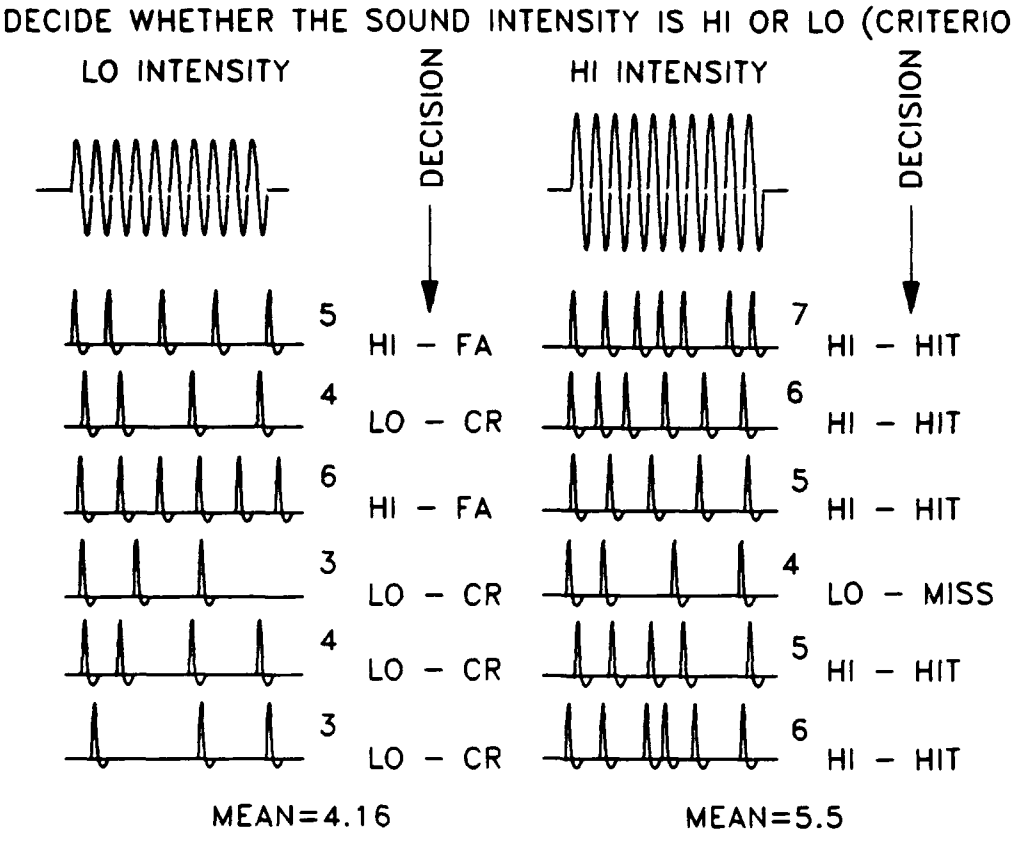


Fig. 7. Idealized responses from a single auditory fiber to multiple repetitions of a "low" and "high" intensity tone burst.

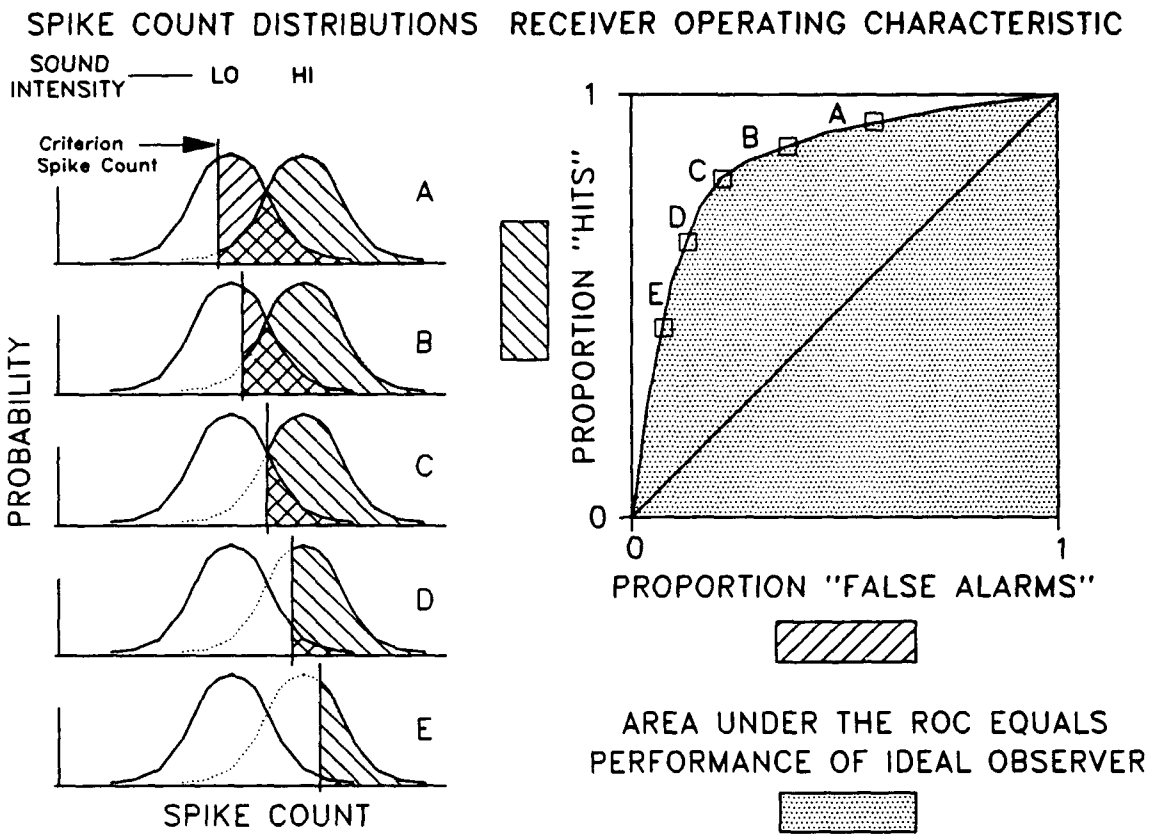


Fig. 8. Idealized spike cound distributions (or pulse number distributions) for two sounds of different intensity. The placing of criteria at different points along the spike count decision axis results in the definition of the ROC function. The area under the ROC estimates optimum performance in a 2IFC task.

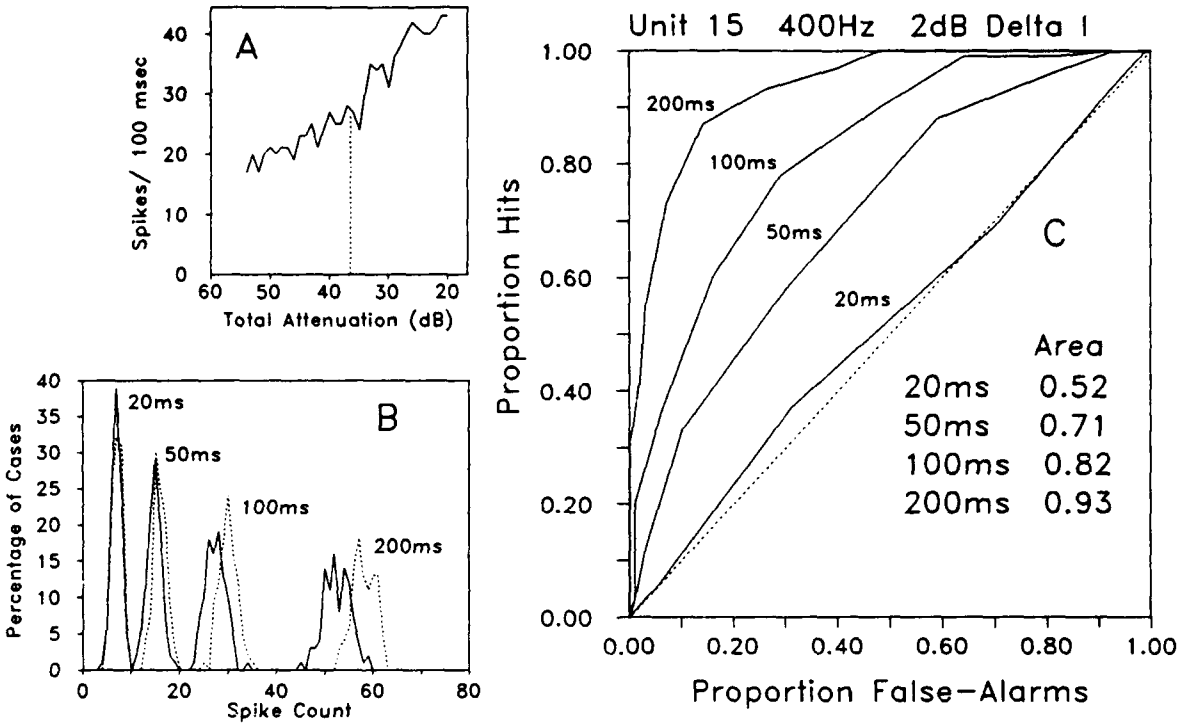


Fig. 9. Steps in conducting the neurophysiological experiment. First, a spike count - sound intensity function is obtained. Then at some level within the fiber's dynamic range, spike count distributions are obtained for two intensity levels at four durations. These define four ROCs (one for each duration), and estimate the optimum performance expected for listening over each of the durations.

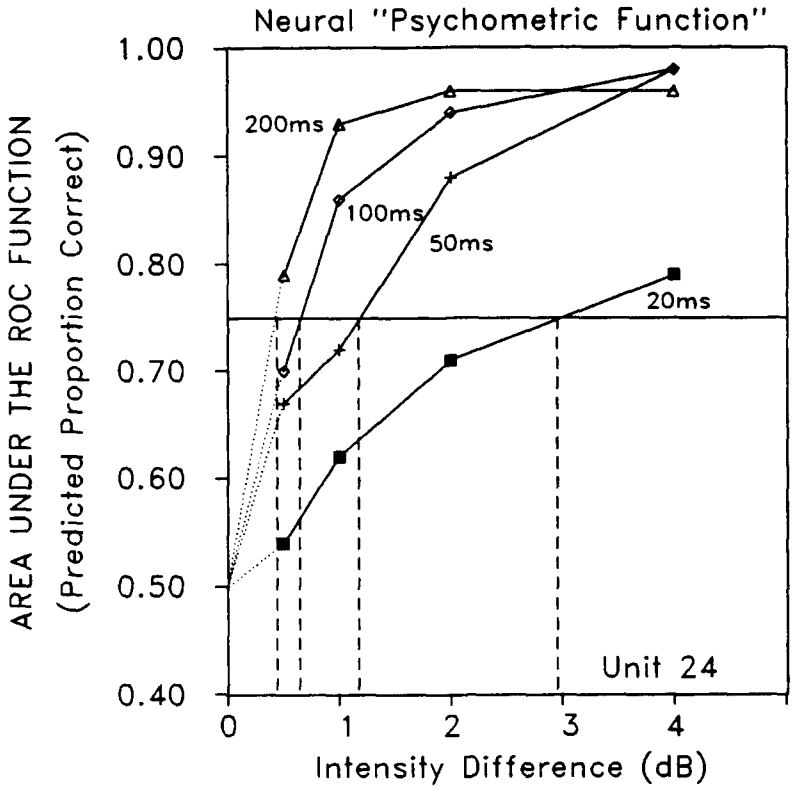


Fig. 10. Neural "psychometric functions" obtained from one auditory nerve fiber of the goldfish at four signal durations.

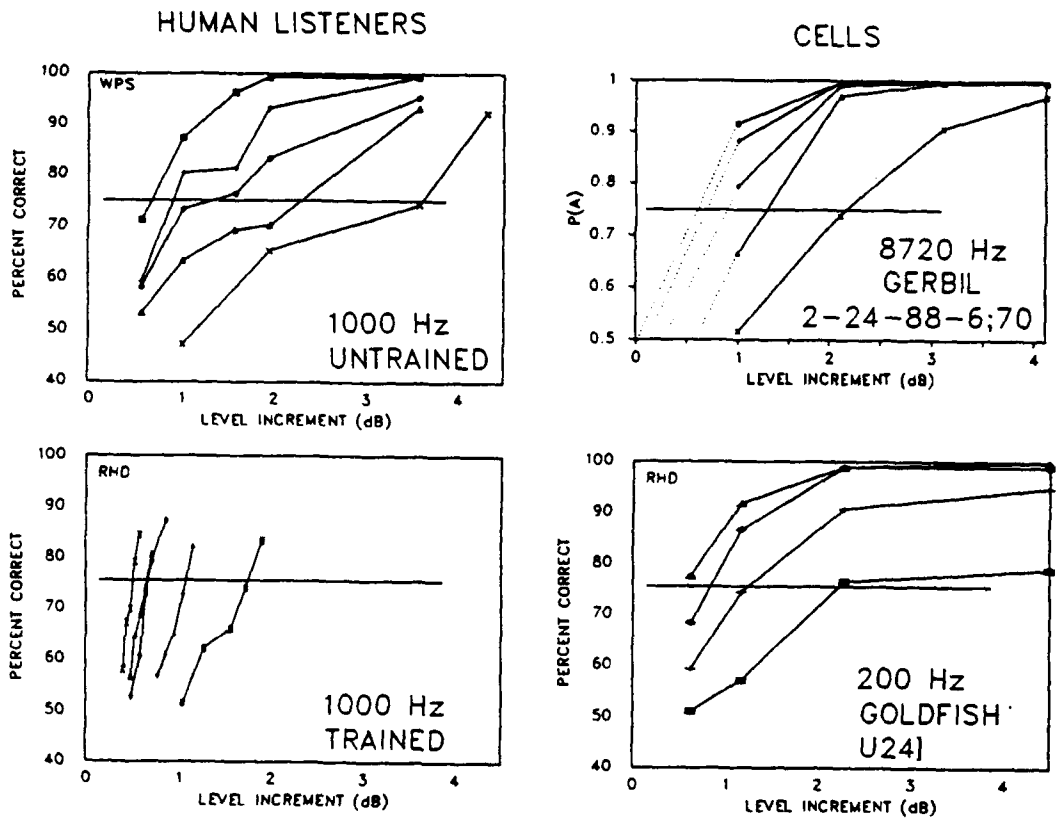


Fig. 11. Psychometric functions at different durations for two human listeners and for one cell from the gerbil cochlear nucleus (chop-s), and one cell from the goldfish auditory nerve. Within each panel, the durations are (from left to right) 20, 50, 100, 200, and, for the human and gerbil, 400 ms.

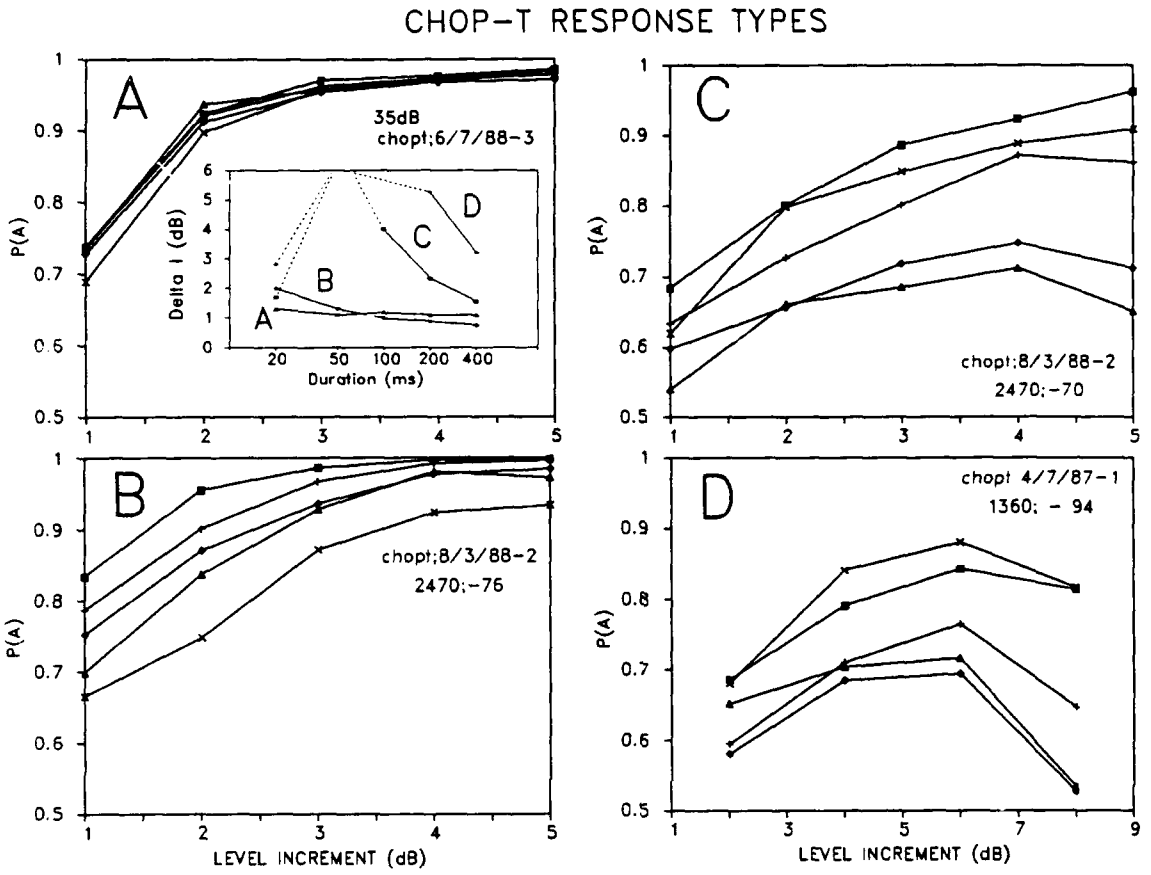


Fig. 12. Neural "psychometric functions" at different durations for three transient chopper cells of the gerbil cochlear nucleus. Note that panels B and C are the same cell tested at different overall intensities (6dB apart). This level difference changes the cell from showing a monotonic duration function (B) to a non-monotonic function (C).

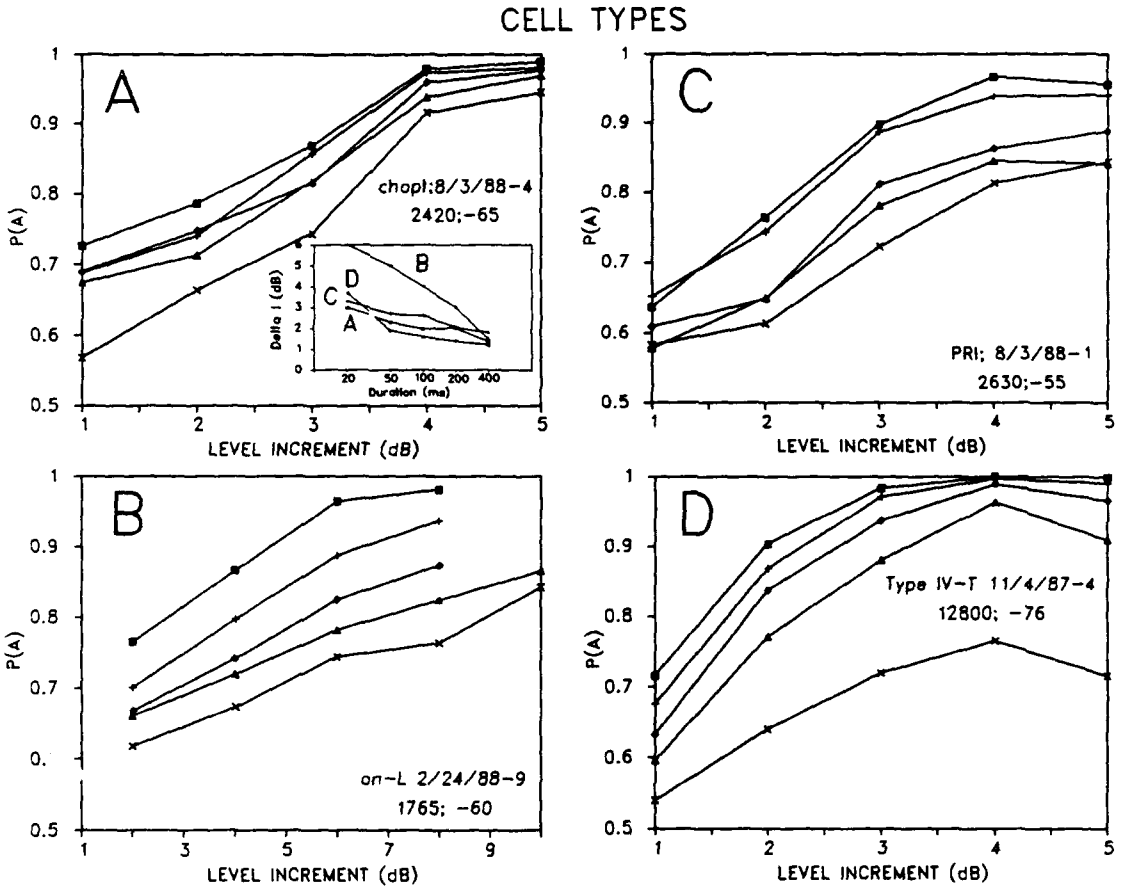


Fig. 13. Neural "psychometric functions" at different durations for four different cell types of the gerbil cochlear nucleus. Cells were chosen to illustrate that different cell types could show similar behavior.

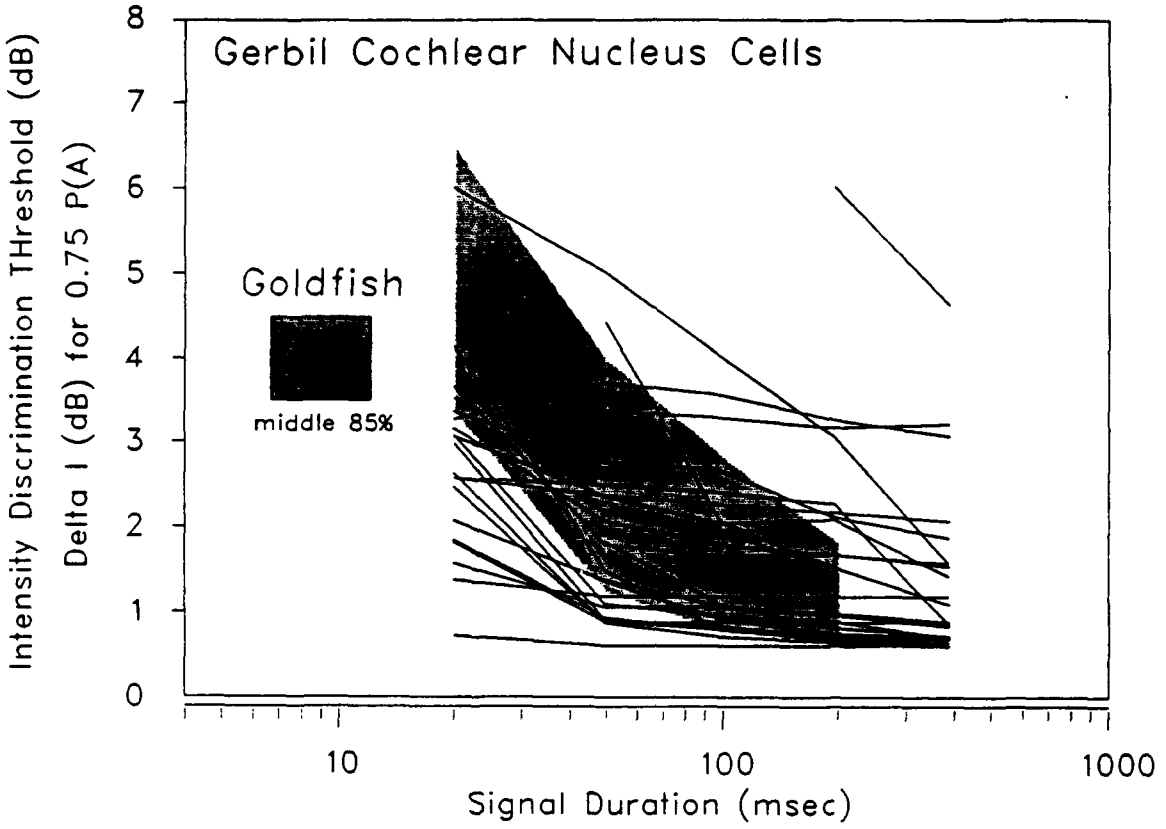


Fig. 14. Duration functions for about 25 cells of the gerbil cochlear nucleus. These show the intensity differences corresponding to 76% correct performance. The functions are shown at the overall intensity producing the lowest thresholds (best performance).

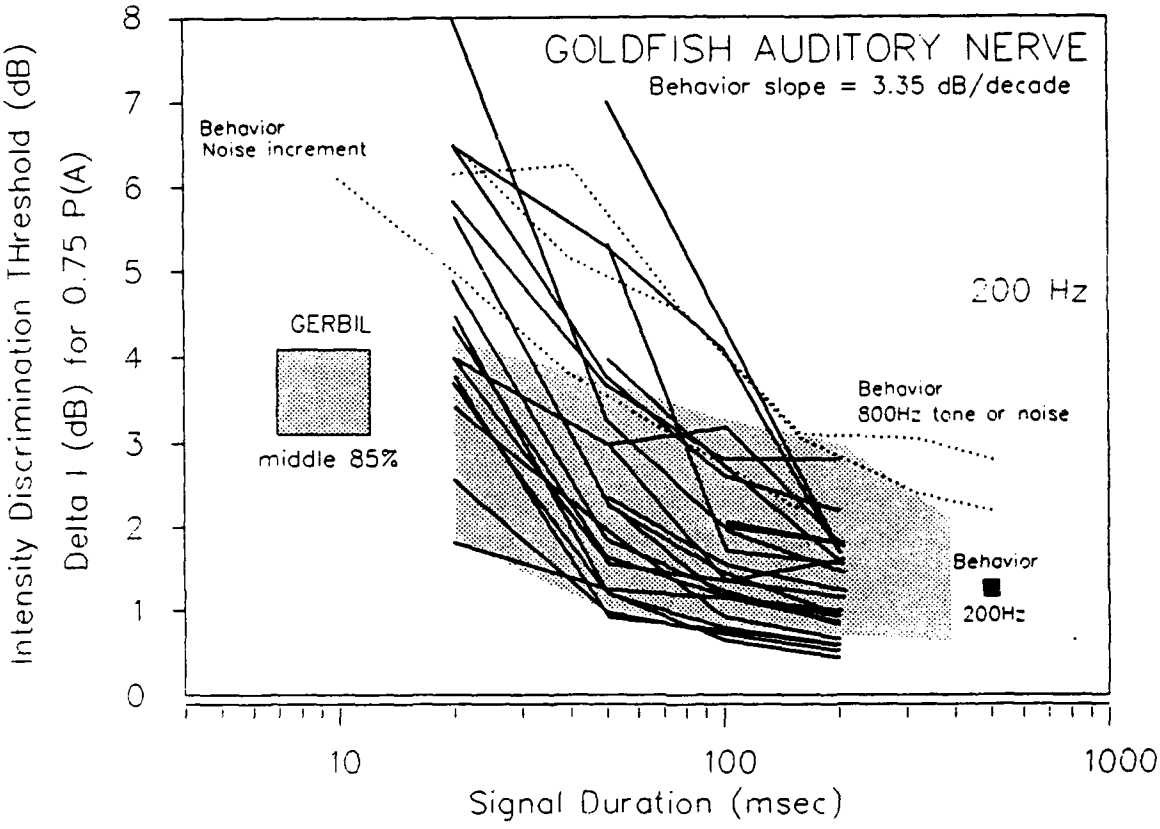


Fig. 15. Duration functions for about 25 cells of the goldfish auditory nerve. These show the intensity differences corresponding to 76% correct performance. The functions are shown at the overall intensity producing the lowest thresholds (best performance).

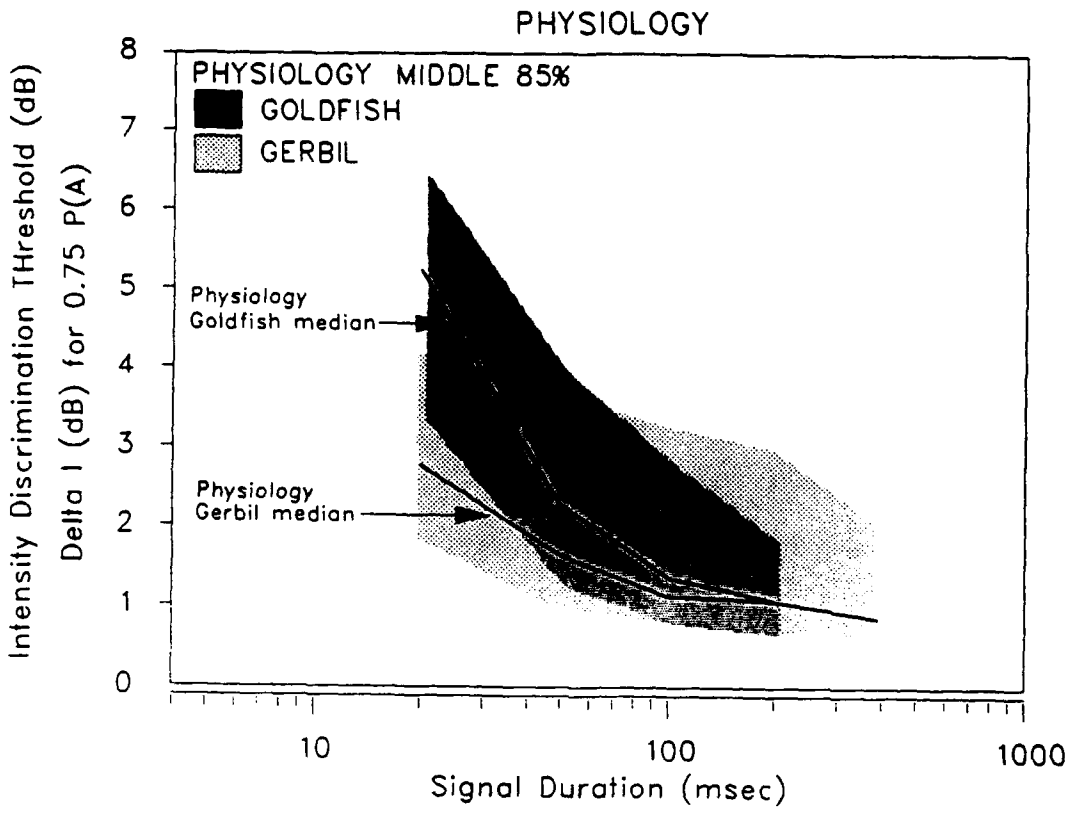


Fig. 16. Shaded areas are the middle 85% of the duration functions of the cells shown in Figs. 14 and 15 for the gerbil and goldfish. Dark lines are medians for each species.

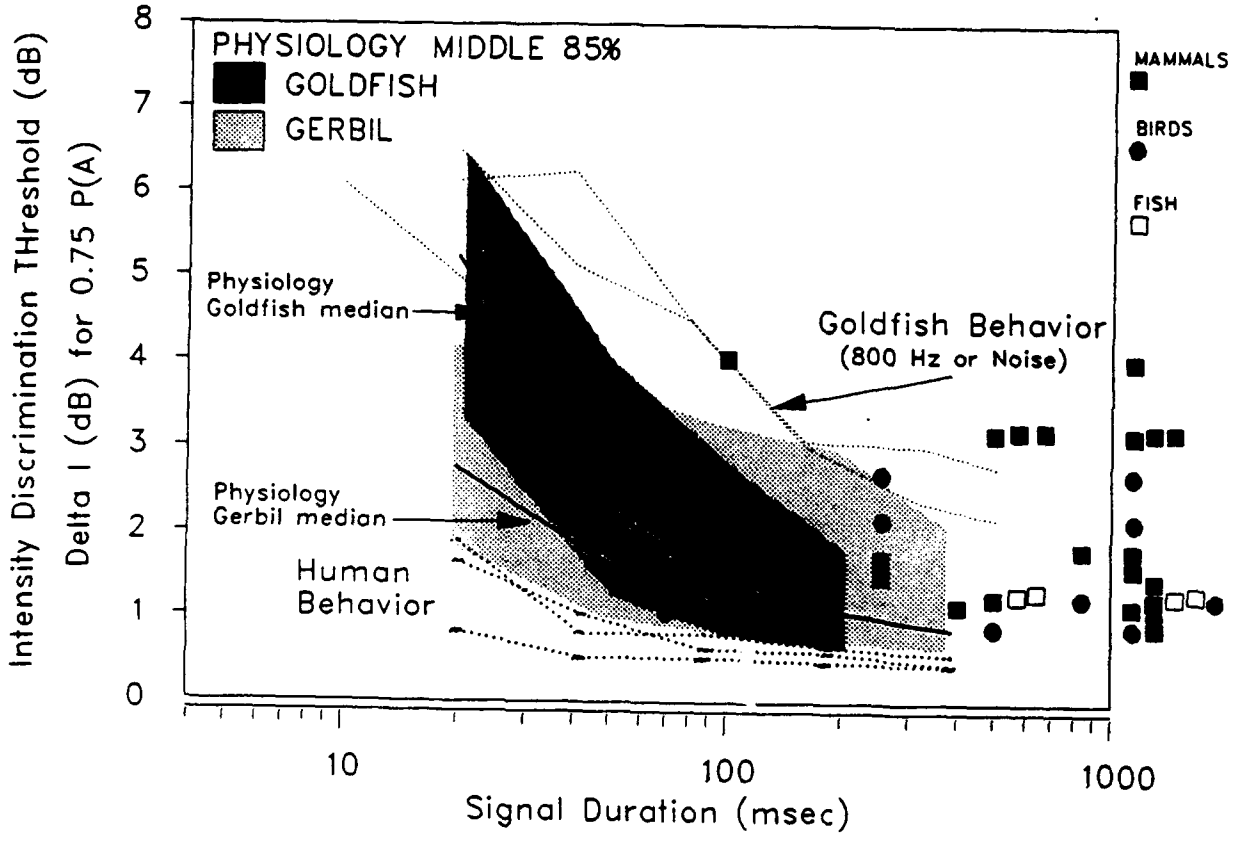


Fig. 17. A comparison of psychophysical results with neurophysiological results on the effect of duration on intensity discrimination performance of human and non-human listeners, and of goldfish auditory nerve fibers and gerbil cochlear nucleus cells. Data are replotted from Figs. 5 and 16.

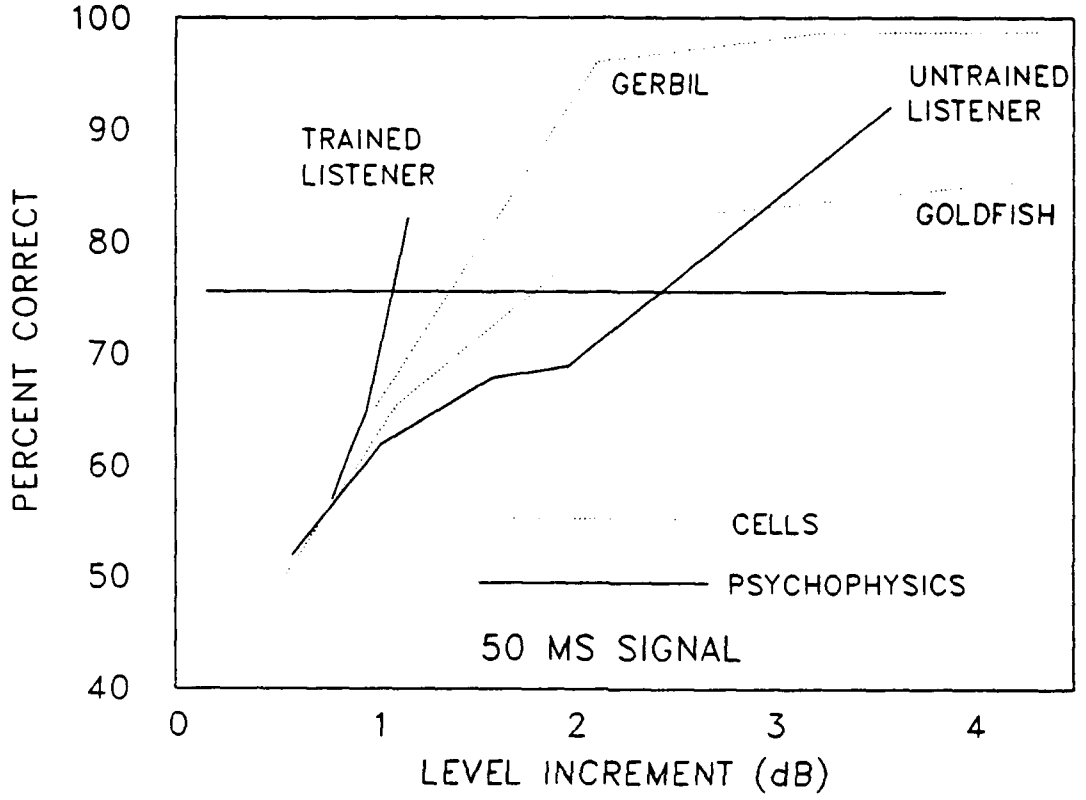


Fig. 18. Psychometric functions for two human observers (one trained and one untrained) discriminating intensity differences between 50ms signals compared with selected cells of the gerbil cochlear nucleus and the goldfish auditory nerve. The differences in human performance may exceed the species and level-of-processing differences observed at the cellular level.

Physiological correlates of forward masking in single nerve-fiber and compound neural responses recorded from the auditory nerve.

Evan M. Relkin[1,2], Christopher W. Turner[3,1], John R. Doucet[1,2] and Robert L. Smith[1,2]
([1]Institute for Sensory Research, [2]Department of Bioengineering, and [3]Communication Sciences and Disorders, Syracuse University, Syracuse, NY 13244-5290)

Introduction

It is often assumed that forward masking is a relatively simple phenomenon that is completely determined by adaptation in the auditory periphery, even though many of those who study adaptation caution against this inference. Perhaps one reason for this assumption is that forward-masked tuning curves for single auditory nerve fibers, or the compound action potential, show a strong similarity to psychophysical tuning curves. However, most, if not all, studies of what was termed forward masking for eighth nerve fibers, or the compound action potential, have measured response reduction, not elevation in threshold. Thus the evidence that forward masking is entirely peripheral is indirect. I hope to show that the story is somewhat more complicated. I will present our methods and data for forward-masked detection thresholds for single fibers and whole nerve responses measured in the chinchilla. We use the whole nerve responses as a convenient measure of the activity of the entire nerve bundle. I will start with the single neuron experiments.

Two-Interval Forced-Choice Procedure

Figure 1 (from Relkin and Pelli, 1987) illustrates the method for measuring the detection threshold for the spike trains recorded in response to test tones . This method was developed in our laboratory (Relkin and Pelli, 1987) and simultaneously, by Delgutte (1988). The method is based on the familiar Two-Interval Forced-Choice (2IFC) procedure. A trial consists of two intervals; the probe is present in one as selected randomly. Spikes are counted for two identical time windows in each interval and the number of counts for each interval are compared. A correct detection of the probe occurs when the number of spikes counted in the interval containing the probe exceeds the number in the other interval. Thus, as will be discussed further, threshold can be defined in terms of the probability of a correct detection. Note that unless I specify otherwise, the masker-probe interval, Δt, equals zero, and the frequency of the masker equals the frequency of the probe. Also, for single fibers, the frequency of both equals the characteristic frequency of the neuron.

One way to measure threshold, from the neurometric function, is demonstrated in the upper left panel of Figure 2 (from Relkin and Pelli, 1987). The neurometric function is analogous to the psychometric function; the probability of a correct detection is plotted as a function of the probe level. In this case there was no masker. The solid line is a best-fit Weibull function. Threshold is defined as the intensity for which the performance (i.e. percent correct detections) equals some prechosen value, often 75 %, along this function. Measuring threshold this way is slow, requiring approximately 15 to 20 minutes to collect enough data points to define the neurometric function adequately.

A quicker alternative is to use an adaptive, up-down tracking procedure to find the intensity for which performance equals the chosen value. We use the PEST (parameter estimation by sequential testing; Taylor and Creelman, 1967) procedure. Thresholds can be determined in less than two minutes. Thus it is possible to measure thresholds for many stimulus conditions during the typical holding-time for an auditory neuron (10 minutes to 1 hour).

While collecting data for neurometric functions, it is possible to construct Pulse-Number Distributions, or PND's, for each intensity level. The PND is a summary of the statistics of the

counts recorded in each of the two intervals. An example of a PND for a high spontaneous rate fiber is shown in the upper right panel of Figure 2. Each point on the PND shows the number of times (out of 50 trials) a given number of spikes was counted. The open symbols are for counts recorded in the interval without the probe and the closed symbols are for counts in the interval containing the probe. The probability of a correct detection at the respective stimulus intensity is given in the upper right corner of the panel.

The middle right panel of Figure 2 shows PND's for a greater intensity at which the the probability of a correct detection was near perfect. Notice, relative to the upper right panel, that the overlap of the distributions decreases as the probability of a correct detection increases.

In Figure 3 (from Relkin and Pelli, 1987) similar neurometric functions and PND's are shown but in this case the spontaneous rate of the neuron was low. Looking at the open symbols on the PND's, which correspond to the interval with no probe, it can be seen that, most often, the number of spikes counted is zero. Thus the presence of a spike is a very reliable indication that the probe was present. In this sense, low spontaneous rate fibers behave like 'perfect' detectors. As we will see, the response properties of low spontaneous rate fibers are relevant to spike counts for the time interval following a forward masker.

Forward Masking of Single Neurons

Figure 4 (from Relkin and Turner, 1988) shows PND's recorded from a high spontaneous rate fiber, with (right panels) and without (left panels) the presence of a forward masker that produced a saturated average firing rate. Notice that for the panels on the right, when a masker is present the PND's are very similar to those recorded for a low spontaneous rate fiber. When there is no probe (open symbols) the spike count is most often zero and the presence of a spike is again a very reliable indication of the presence of the probe. In other words, following a masker, due to the adaptation of spontaneous firing rate, there is less noise. Since both the response and the noise are reduced, one could anticipate that the threshold shift is limited.

Figure 5 (from Relkin and Turner, 1988) demonstrates how we define threshold shift. There are two neurometric functions, one recorded without (open symbols) a masker, and the other (closed symbols) recorded with a forward masker that produced a saturated average firing rate. Note that the presence of the masker results in a parallel shift to the right along the intensity axis. Threshold shift is defined as the distance, relative to the intensity axis, between the two best-fit functions. Because the shift of the functions is parallel, the measured threshold shift is independent of the performance level at which the threshold shift is determined. Thus we can use the adaptive procedure to measure the threshold at one performance level, with and without the masker, and measure the threshold shift as the former minus the latter.

The adaptive procedure was used to measure forward masked thresholds as a function of masker intensity to form growth-of-masking functions. Figure 6 (from Relkin and Turner, 1988) shows typical results for a low and a high spontaneous rate fiber. Note that the growth of thresholds slows at high intensities and seems to saturate. Thus the maximum threshold seems to be limited relative to what would be expected for psychophysical forward masking.

The left-most point on each function in Figure 6 was recorded for a masker intensity that produced no increase in firing rate, and thus can be taken as the threshold in quiet. Our maximum masker intensity was always 80 dB SPL. We define the maximum threshold shift as the difference between the maximum threshold and the threshold for the left-most point. Maximum threshold shifts for many neurons are plotted in Figure 7 (from Relkin and Turner, 1988) as a function of fiber spontaneous rate. Note that there is a tendency for the maximum threshold shift to decrease with increasing spontaneous rate. More importantly, the maximum threshold shift is never much greater than 20 dB and for some neurons can be as little as 3 to 5 dB.

In Figure 8, the growth-of-masking function for the high spontaneous rate neuron of Figure 6 is compared to the linear function used by Jesteadt, Bacon, and Lehman (1982) to summarize human psychophysical growth-of-masking functions for nearly identical stimulus conditions. For the psychophysical data, Δt was 5 ms, the minimum value studied by Jesteadt et al. Since the slope of growth-of-masking functions increases as Δt decreases, the discrepancy between the two functions would only be worse if psychophysical data were available for Δt equal to zero. Figure 8 demonstrates that at least in the intensity domain, there is a large discrepancy between thresholds for single neurons and the human observer. Particularly at high masker intensities, the presence of the probe can be detected from the neural spike train at intensities up to 30 dB lower than that required by the human observer. Therefore, it can be concluded that the central nervous system is making suboptimal use of information available from single fibers of CF equal to the probe frequency.

Hypothetical Explanations for the Single Neuron Results

We are considering several hypotheses that might explain this discrepancy. We are in the process of testing these hypotheses and cannot yet offer the final answer.

A first hypothesis is that the listener may be constrained to attend to low spontaneous rate fibers for high masker intensities. This possibility follows from the suggestions of several other researchers that the low spontaneous rate fibers mediate intensity coding at high stimulus levels. For the interval following a high level forward masker, the listener may not be able to instantaneously switch back to the high spontaneous rate fibers when detecting the probe. It is possible that the total growth of masking function results from splicing the low level portion of the function for high spontaneous rate fibers to the function for low spontaneous rate fibers near the intensity at which the former saturates (consider the two functions in Figure 6).

We have indirect evidence that this hypothesis may be rejected. This evidence is based on the dependence of forward masked thresholds for nerve fibers on the masker-probe interval, Δt. Figures 9 and 10 show this dependence for a high spontaneous rate fiber and a low spontaneous rate fiber, respectively. For human psychophysics, masked threshold recovers completely by 80-100 msec following the masker. This is similar to the case for high spontaneous rate fibers. However, low spontaneous rate fibers require up to 300 msec for thresholds to recover completely. Thus, the time course for recovery from forward masking for the low spontaneous rate fibers does not match the psychophysics. Therefore, it seems unlikely that the low spontaneous rate fibers are mediating the detection of forward-masked probe tones.

A second hypothesis is that the observer uses a less optimal temporal processing window. The counting window used in our 2IFC procedure assumes exact knowledge of when the probe will occur and the probe's duration. Basing detection of the probe on counts in other windows, or perhaps integrating over some exponential temporal window, could degrade detection. Another way of stating this hypothesis is that the observer is unsure when the probe will occur.

A third, complementary, hypothesis is that performance is degraded by spatial summation of the responses of fibers along the cochlear partition. In other words, detection is not based only on fibers with CF corresponding to the probe frequency. Perhaps the observer is constrained to base detection of the probe on the responses of all fibers that respond to the masker. Off-CF fibers that do not respond to the probe would only add noise to the detection process without adding any useful information, thereby increasing thresholds. Another way of stating this hypothesis is that the observer is unsure of the frequency of the probe.

Forward Masking of CAP

We use the compound action potential as a simple measure of the net output of the auditory nerve to test the effects of spatial summation. A population study of individual fibers would be

more direct but these experiments are difficult and tedious and interfere with normal sleeping habits. Parenthetically, we have found that the CAP has remarkable ability to predict behavioral performance on other tasks, such as measuring the intensity difference limen for clicks.

The 2IFC procedure that we have developed to measure detection thresholds for the CAP is illustrated in Figure 11. The method is identical to that used for single fibers with the exception that the decision variable in each interval is the magnitude of the most negative peak, which corresponds to N_1 of the CAP. A correct detection occurs when the larger negative peak occurs in the interval for which the probe was present. Using this definition of a correct detection we can measure neurometric functions and find thresholds using the adaptive procedure as described earlier for single fibers.

We have measured growth-of-masking functions for the CAP for a masker and probe frequency of 5 kHz. Results are shown in Figure 12 for four values of the masker-probe interval. These growth-of-masking functions do not saturate and can be summarized by best-fit linear functions as shown in the figure. The slope for Δt equal to 2 ms is close to the value found for human psychophysics. In addition the slope decreases with increasing Δt as is also seen for human psychophysics. The dependence of threshold on Δt is shown in Figure 13. When Δt is plotted on a logarithmic axis, the data can again be fit by linear functions. The two linear dependencies shown in Figures 12 and 13 are functionally the same as found by Jesteadt et al. for human observers. In Figure 14, the growth-of-masking function for the CAP at Δt equal to 2 ms is added to the comparison between physiology and psychophysics that was made in Figure 8. The similarity between forward masking of the CAP and the human psychophysics is clear.

We have been considering the significance of the correspondence between forward masking of the CAP and psychophysical forward masking. One possibility is that the agreement is serendipitous, although, as mentioned above, other studies using the CAP now seem to argue against this possibility. Since the CAP is a population response, these results seem to support the importance of spatial summation. However, the CAP is also an onset response and it is possible that the onset response has a weighted role in the detection of short probe tones following a masker. This last possibility is suggested by PST histograms of the kind shown in Figure 15(from Relkin and Turner, 1988). If you compare the response to the probe tone with and without the masker, it can be seen the onset response is reduced by a greater proportion than the response at the end of the probe tone. This *suggests* that the onset may be determining forward masked thresholds. However, without considering the variance of the response throughout the duration of the probe tone it is not possible to infer threshold shifts from reductions in the PST's. We intend to test the importance of the onset response by modifying the 2IFC procedure for single fibers by using counting windows that give added weight to early responses.

Peristimulus Compound Action Potential

It would also be possible to investigate the relative importance of the onset response if it were possible to record compound neural activity during the steady-state response to a tone burst. This has not been possible with all former methods for recording the CAP. In collaboration with John Doucet, we have developed a method for recording from the auditory nerve at the internal meatus with a gross suction electrode that makes possible for the first time, the recording of steady-state compound potentials. The basis for this technique is that the contribution of individual neurons to the compound response is monophasic so that synchronization of neural firings is not required to produce a net potential. For conventional recording of the CAP, the contribution of individual neurons is biphasic so that without synchronization, responses sum to zero. Unfortunately, I do not have enough time to provide further details on our new method.

An example of a signal-averaged response to a 300 msec tone burst at 5 kHz at 50 dB SPL is

shown in Figure 16. We have named this response the peristimulus compound action potential (PCAP). It is evident that there is indeed a steady-state response. The similarity to auditory nerve fiber PST's is striking. We are currently investigating the properties of the PCAP to test its relationship to the net firing rate of the auditory nerve.

Figure 17 shows an input/output function for the steady-state portion of the PCAP recorded in response to a 100ms tone burst at 1 kHz. The function has been fit with two linear segments. Since the axes are log/log the linear relationships imply a power law. At low intensities the best fit slope, or equivalently the exponent of the power law, is 0.74 dB/dB. Because of the noise in the data at low intensities, depending on just which points were fit, this slope could have been anywhere between 0.7 and 1.4 dB/dB. At high intensities the slope is clearly 0.31 dB/dB. These values are very similar to slopes that fit growth-of-loudness functions and are also in good agreement with calculations of the total neural firing rate from a model developed by J.L. Goldstein (1974). In addition, my own calculations of total firing rate based on the population studies of Kim and Molnar (1979), taking into account neural density, show a similar dependence on intensity. Unfortunately, the population studies were only done at three intensities. We intend to extend these observations. In summary, the evidence examined to date supports the conclusion that the magnitude of the PCAP is proportional to the net firing rate of the entire auditory nerve.

We have not yet measured forward-masked detection thresholds for the PCAP. However, in Figure 18 we show some examples of response reduction. Masker and probe frequency were both 5 kHz. Moving down the figure, masker intensity was increased while the probe level was held constant. It is clear that there is a greater decrease in the onset response to the probe compared to the later response (not truly steady-state since the probe duration was 25 ms). Again, without studying the response variability it cannot yet be concluded that the onset response mediates forward-masked thresholds. We intend to measure detection thresholds for the onset and late response of the PCAP to address this issue directly.

(Supported by the Whitaker Foundation, the Deafness Research Foundation, and the National Institutes of Health)

References (Since this is a transcript of a talk, these references are not intended to be complete or even representative)

Delgutte , B. (1988). "Physiological mechanisms of masking.' in Basic Issues in Hearing, edited by H. Duifhuis, J.W. Horst, and H.P. Wit (Academic Press, London). pp. 204-214.

Goldstein, J.L. (1974). "Is the power law simply related to the driven spike response rate from the whole auditory nerve?" in Sensation and Measurement, edited by H.R. Moskowitz, B. Scahrf, and J.C. Stevens (D. Reidel, Dordrecht). pp. 223-229.

Jesteadt, W., Bacon, S.P., and Lehman, J.R. (1982). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. AM. 69, 1735-1745.

Kim, D.O. and Molnar, C.E. (1979). "A population study of cochlear nerve fibers: Comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones." J. Neurophysiol. 42, 16-30.

Relkin, E.M. and Pelli, D.G. (1987). "Probe tone thresholds in the auditory nerve measured by two-interval forced-choice procedures." J. Acoust. Soc. Am. 82, 1679-1691.

Relkin, E.M. and Turner, C.W. (1988). "A reexamination of forward masking in the auditory nerve." J. Acoust. Soc. Am. 84, 584-591.

Taylor, M.M. and Creelman, C.D. (1967). "PEST: Efficient estimates on probability functions." J. Acoust. Soc. Am. 41, 782-787.

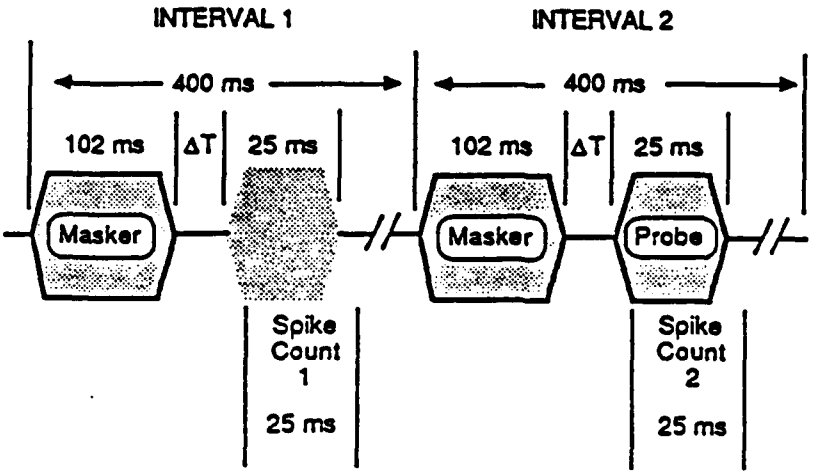

INTERVAL 1
INTERVAL 2
400 ms
400 ms
102 ms
ΔT
25 ms
102 ms
ΔT
25 ms
Masker
Masker
Probe
Spike Count 1
25 ms
Spike Count 2
25 ms

FIGURE 1

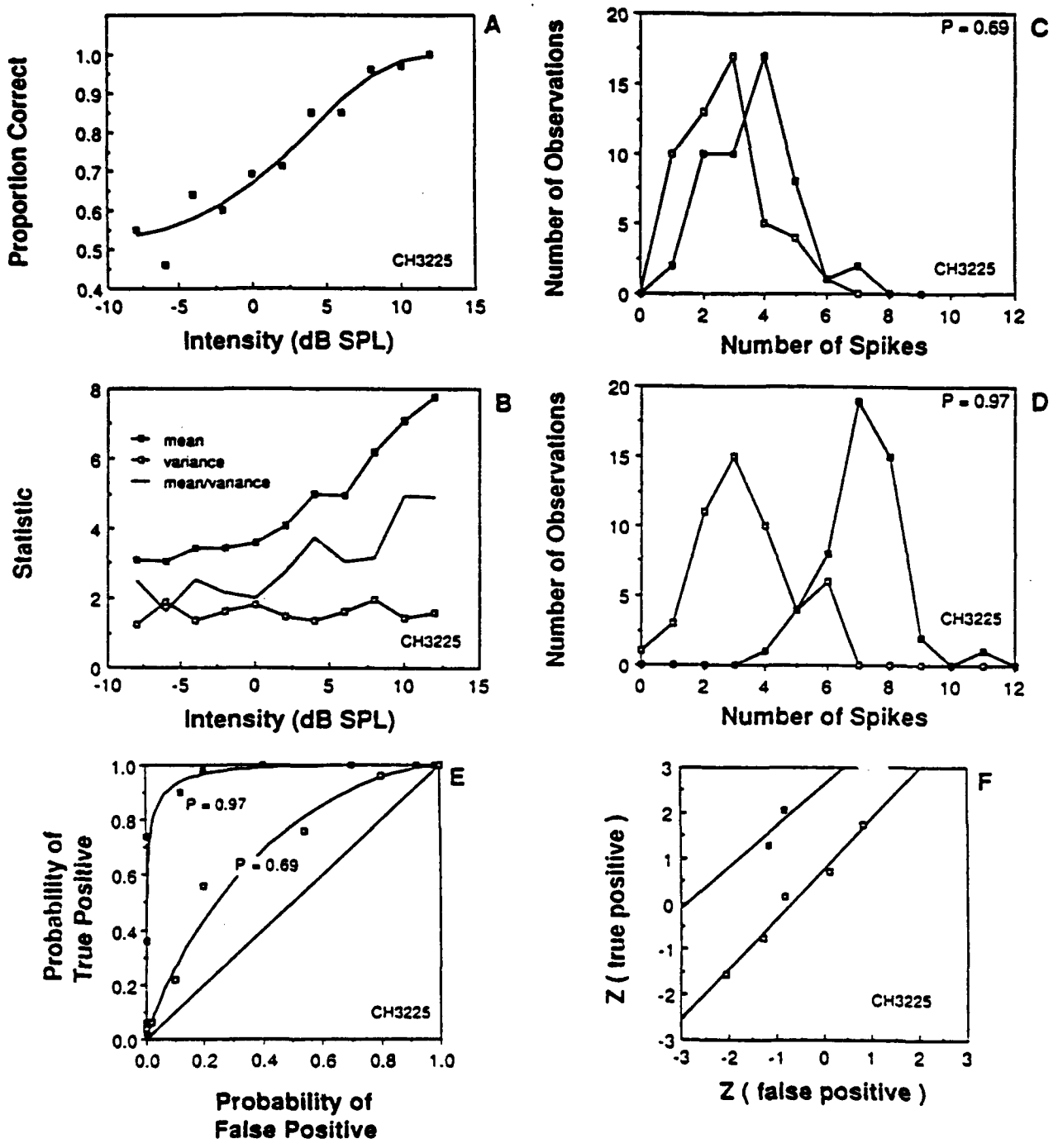

A
Proportion Correct
Intensity (dB SPL)
CH3225
B
Statistic
mean
variance
mean/variance
Intensity (dB SPL)
CH3225
C
Number of Observations
P = 0.69
Number of Spikes
CH3225
D
Number of Observations
P = 0.97
Number of Spikes
CH3225
E
Probability of True Positive
P = 0.97
P = 0.69
Probability of False Positive
CH3225
F
Z (true positive)
Z (false positive)
CH3225

FIGURE 2

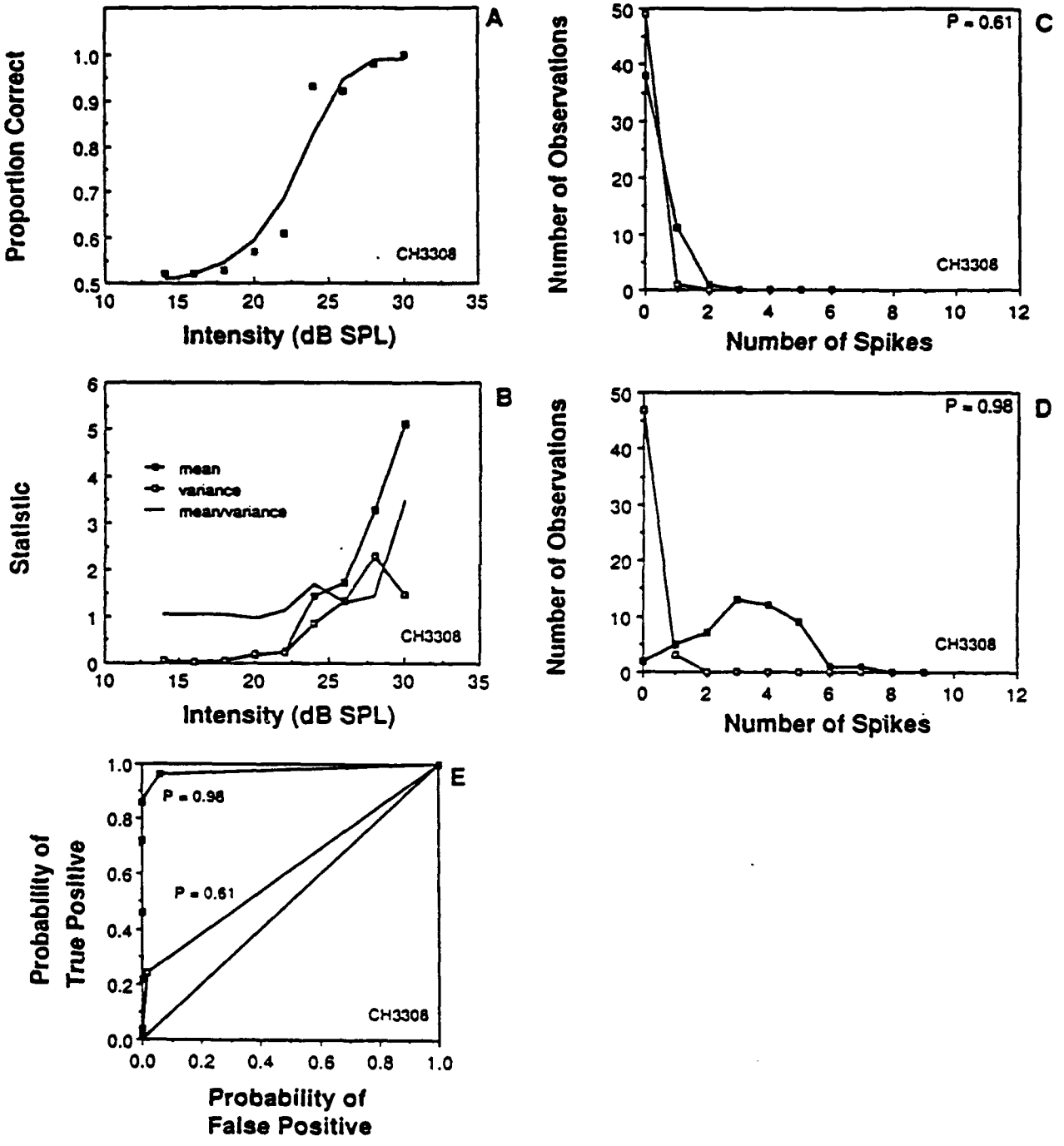
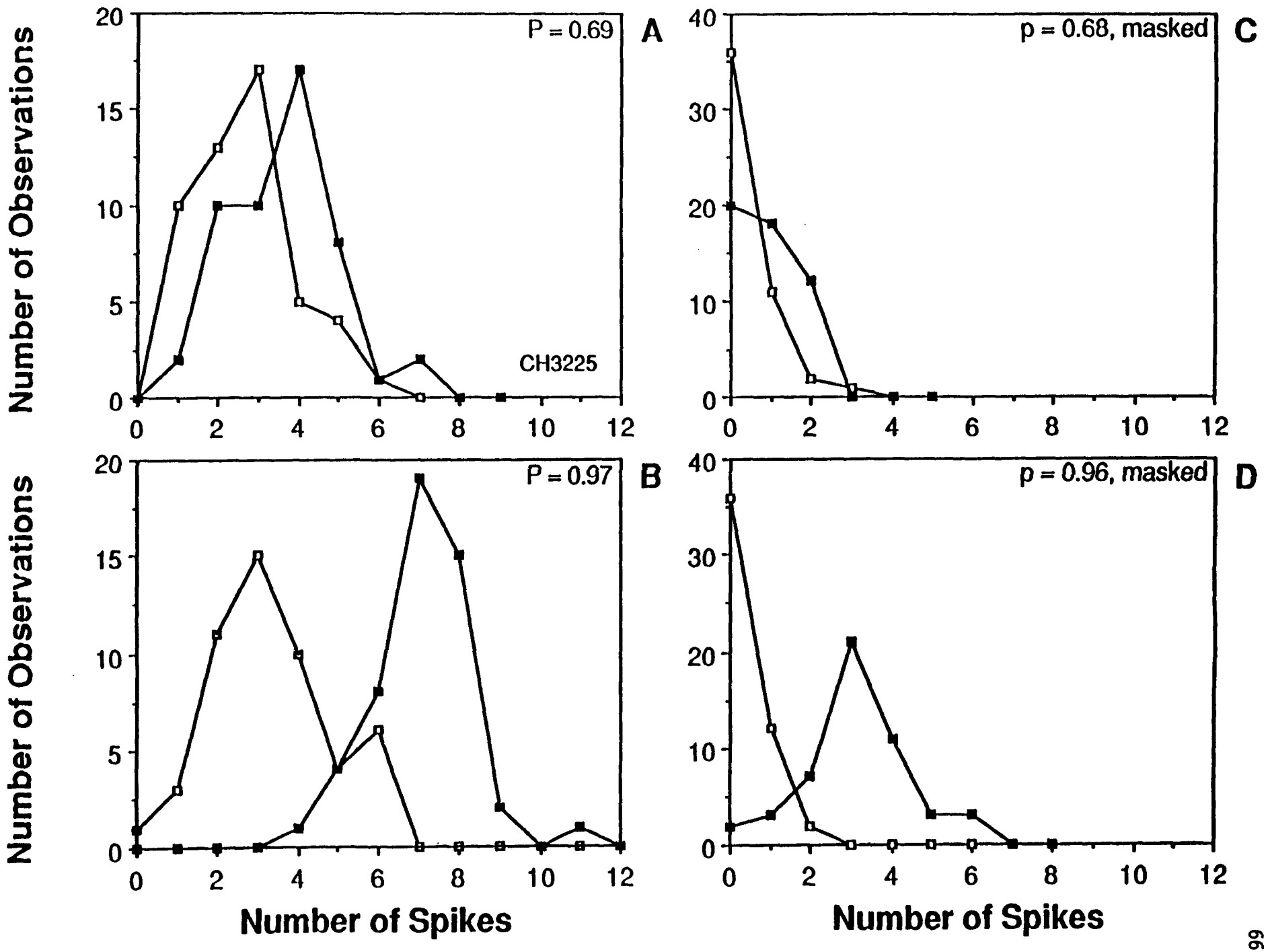

A
Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
10
15
20
25
30
35
Intensity (dB SPL)
CH3308
B
Statistic
6
5
4
3
2
1
0
mean
variance
mean/variance
Intensity (dB SPL)
CH3308
C
Number of Observations
50
40
30
20
10
0
P = 0.61
0
2
4
6
8
10
12
Number of Spikes
CH3308
D
Number of Observations
P = 0.98
Number of Spikes
CH3308
E
Probability of True Positive
1.0
0.8
0.6
0.4
0.2
0.0
P = 0.98
P = 0.61
0.0
0.2
0.4
0.6
0.8
1.0
Probability of False Positive
CH3308

FIGURE 3


Number of Observations
P = 0.69
A
CH3225
p = 0.68, masked
C
P = 0.97
B
p = 0.96, masked
D
Number of Spikes
Number of Spikes

99

FIGURE 4

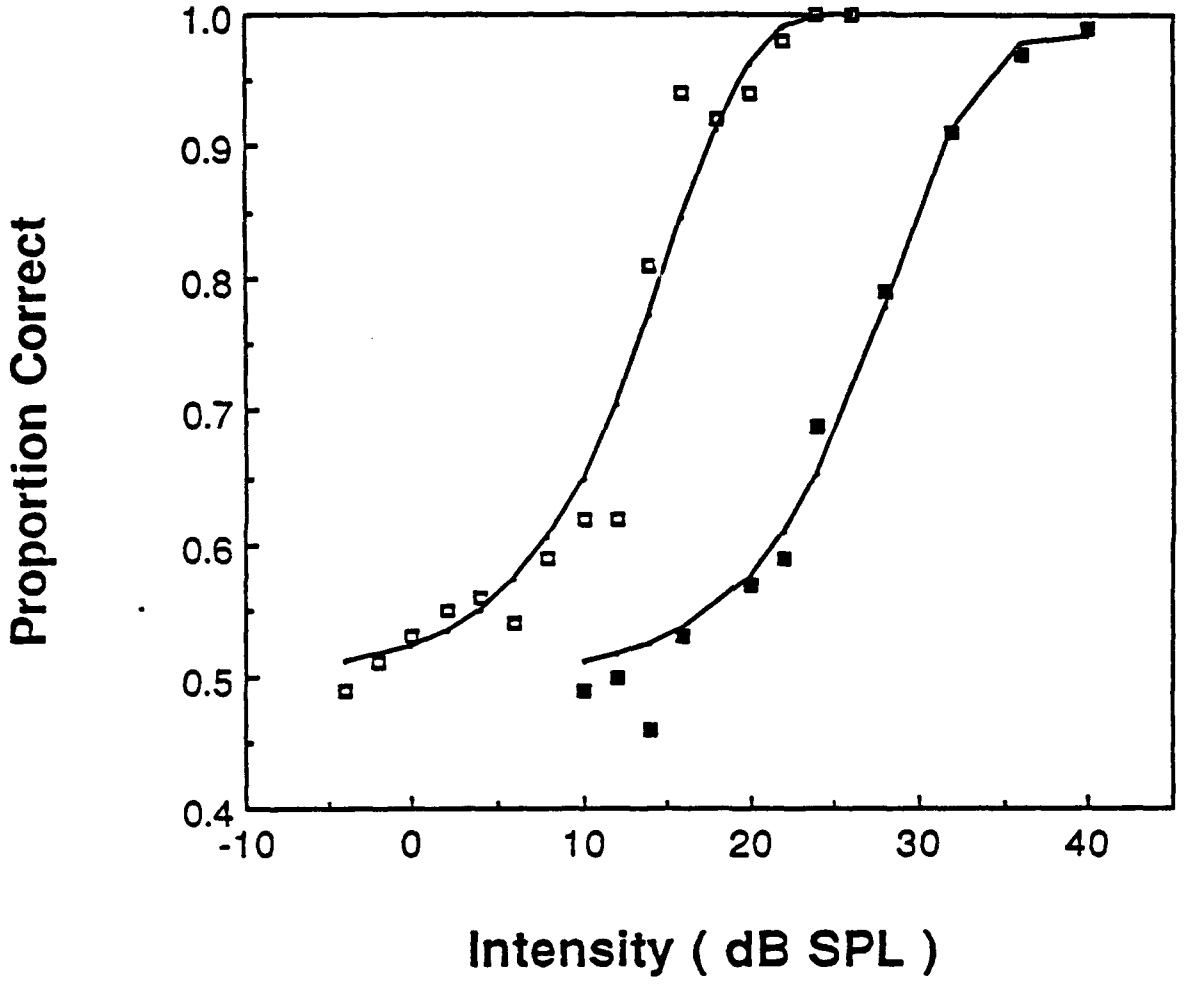

Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
0.4
-10
0
10
20
30
40
Intensity (dB SPL)

FIGURE 5

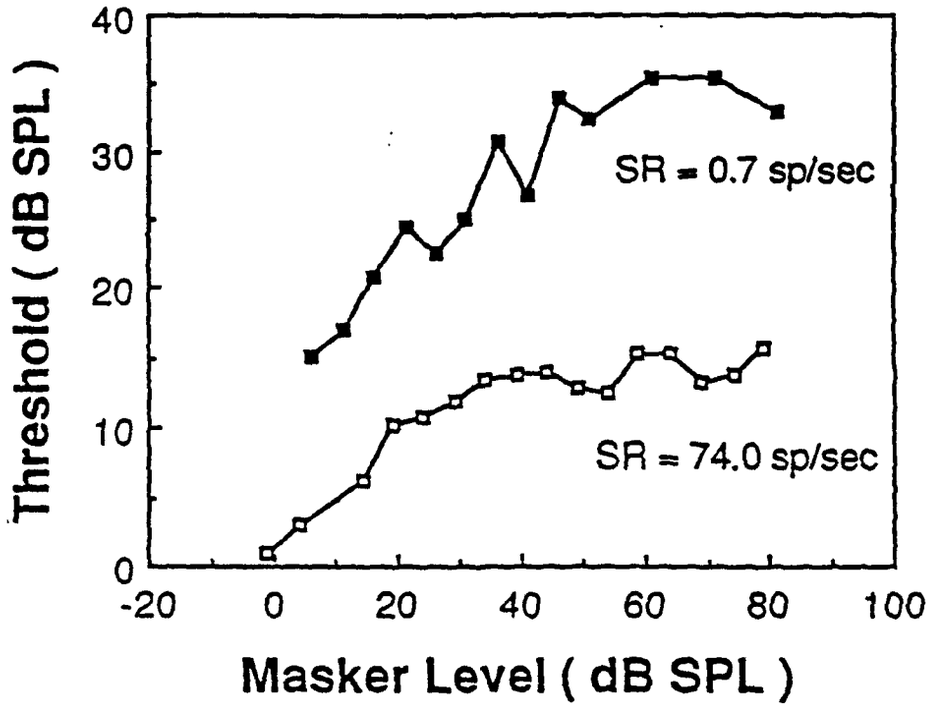

Threshold (dB SPL)
40
30
20
10
0
-20
0
20
40
60
80
100
Masker Level (dB SPL)
SR = 0.7 sp/sec
SR = 74.0 sp/sec

FIGURE 6

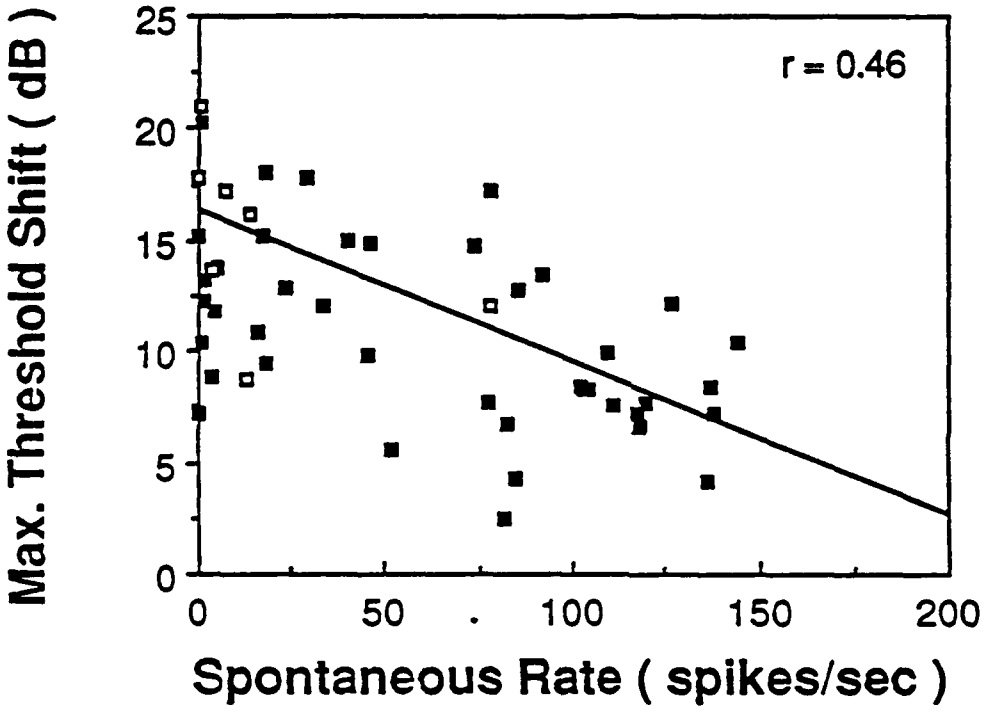

Max. Threshold Shift (dB)
25
20
15
10
5
0
r = 0.46
0
50
100
150
200
Spontaneous Rate (spikes/sec)

FIGURE 7

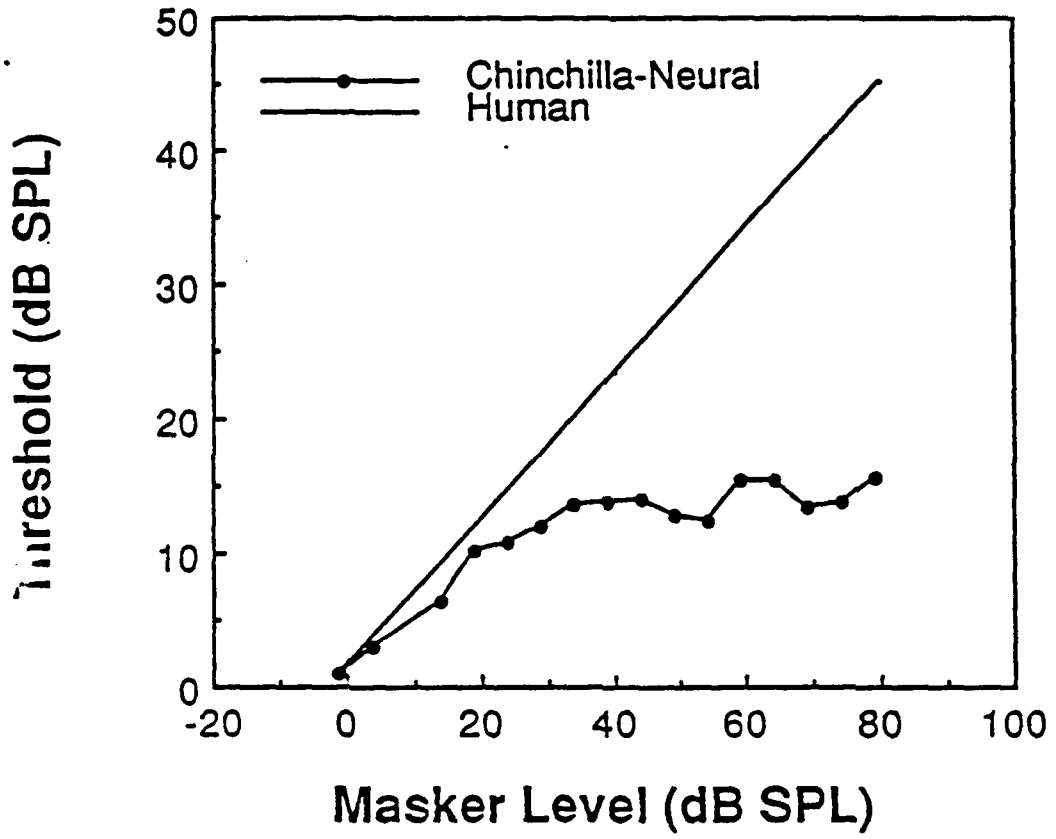

50
40
30
20
10
0
-20
0
20
40
60
80
100
Chinchilla-Neural
Human
Threshold (dB SPL)
Masker Level (dB SPL)

FIGURE 8

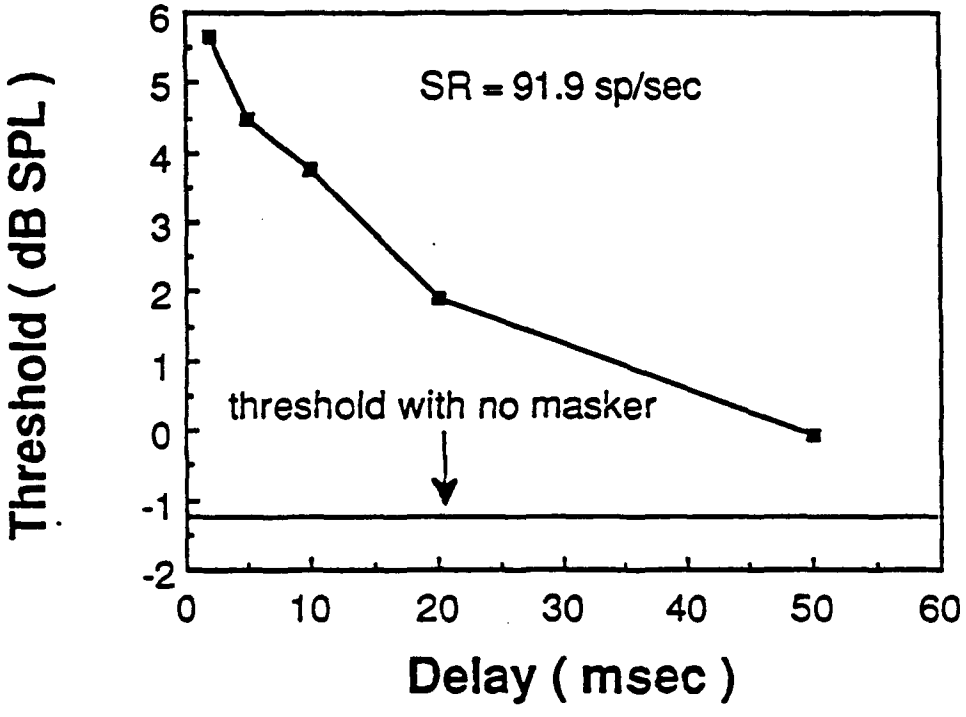

Threshold (dB SPL)
6
5
4
3
2
1
0
-1
-2
SR = 91.9 sp/sec
threshold with no masker
0
10
20
30
40
50
60
Delay (msec)

FIGURE 9

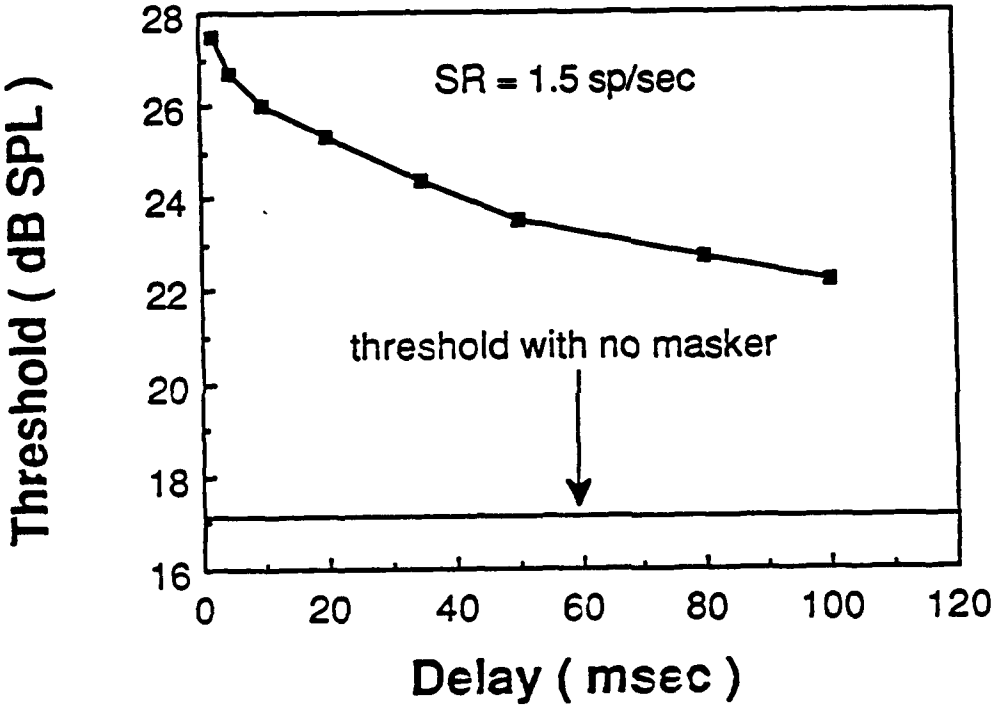

Threshold (dB SPL)
28
26
24
22
20
18
16
SR = 1.5 sp/sec
threshold with no masker
0
20
40
60
80
100
120
Delay (msec)

FIGURE 10

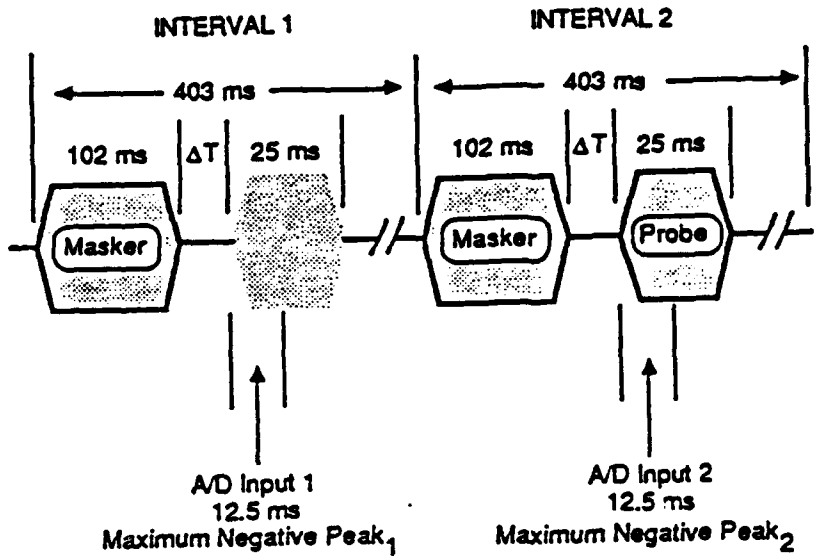

INTERVAL 1
INTERVAL 2
403 ms
403 ms
102 ms
ΔT
25 ms
102 ms
ΔT
25 ms
Masker
Masker
Probe
A/D Input 1
12.5 ms
Maximum Negative Peak1
A/D Input 2
12.5 ms
Maximum Negative Peak2

FIGURE 11

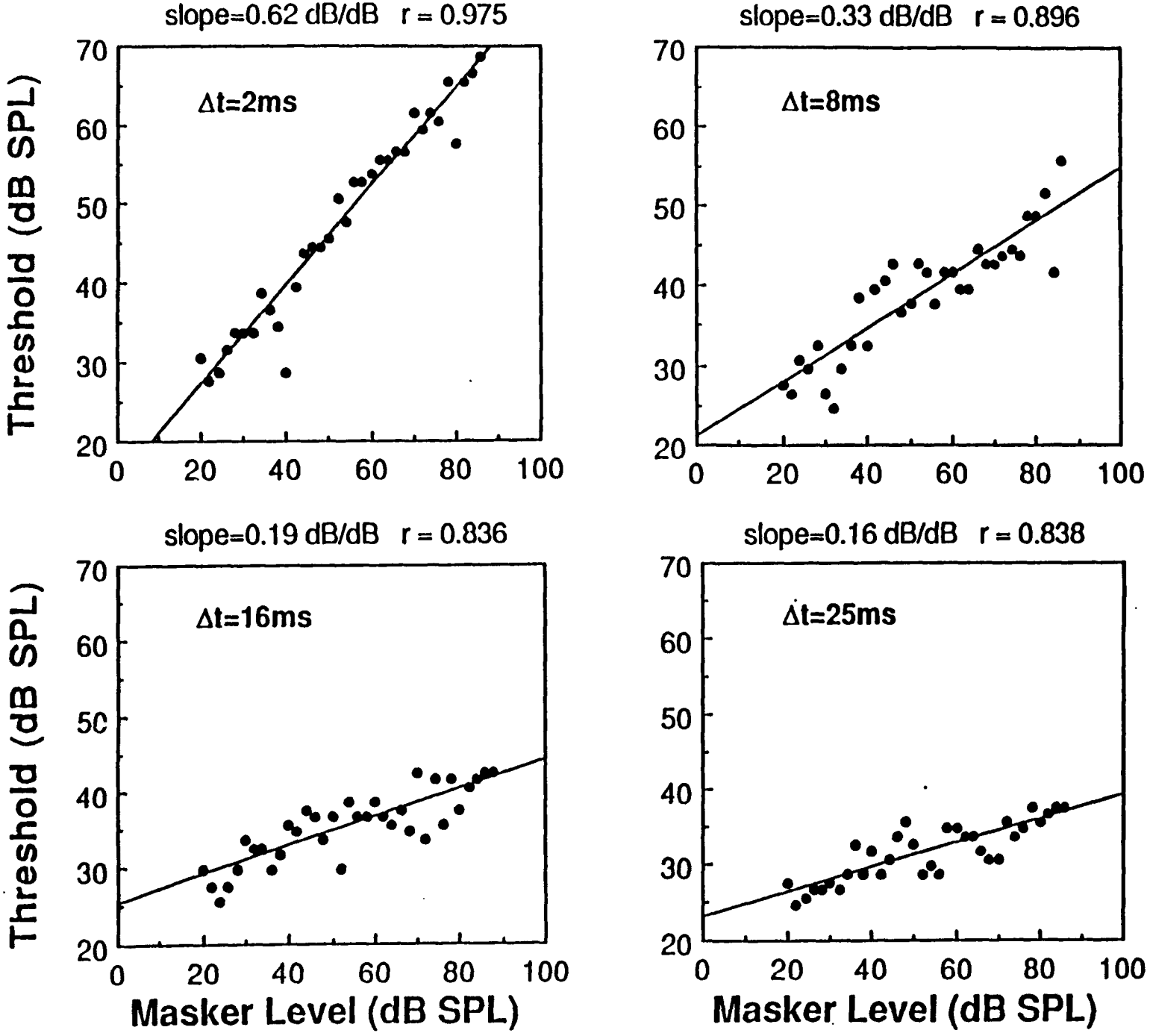

slope=0.62 dB/dB r = 0.975
Δt=2ms
slope=0.33 dB/dB r = 0.896
Δt=8ms
slope=0.19 dB/dB r = 0.836
Δt=16ms
slope=0.16 dB/dB r = 0.838
Δt=25ms
Threshold (dB SPL)
Masker Level (dB SPL)
20
30
40
50
60
70
0
20
40
60
80
100

Figure 12

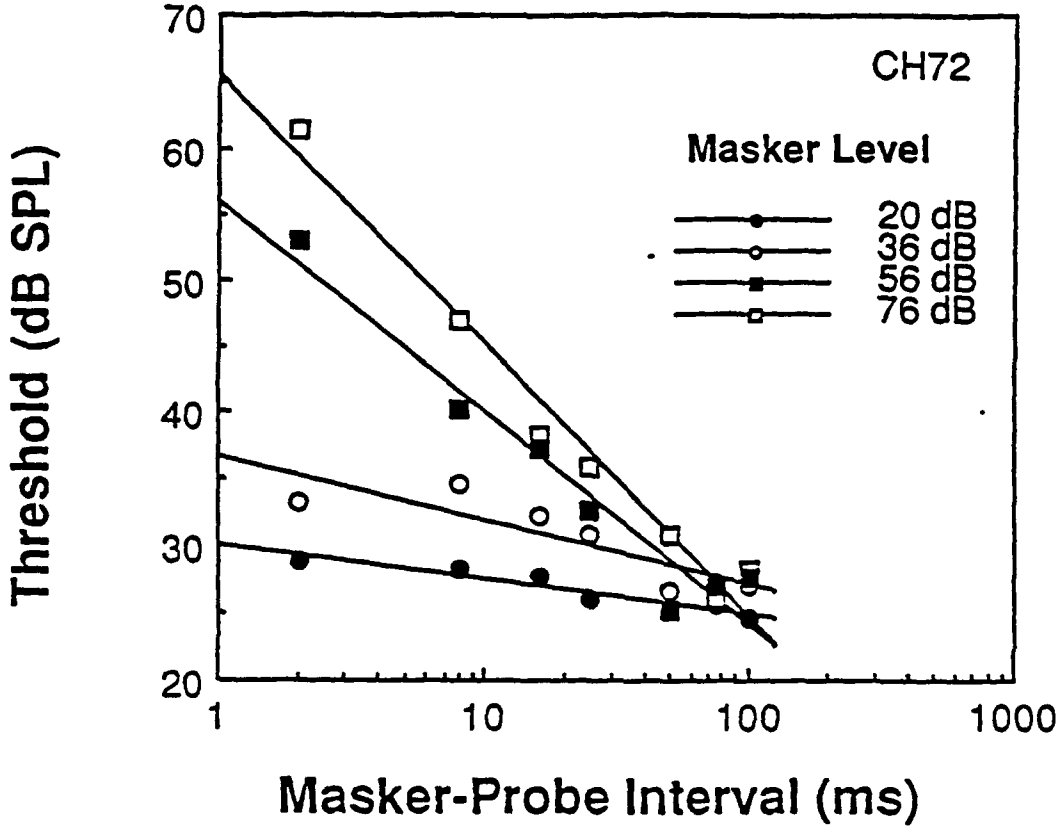

70
60
50
40
30
20
Threshold (dB SPL)
CH72
Masker Level
20 dB
36 dB
56 dB
76 dB
1
10
100
1000
Masker-Probe Interval (ms)

FIGURE 13

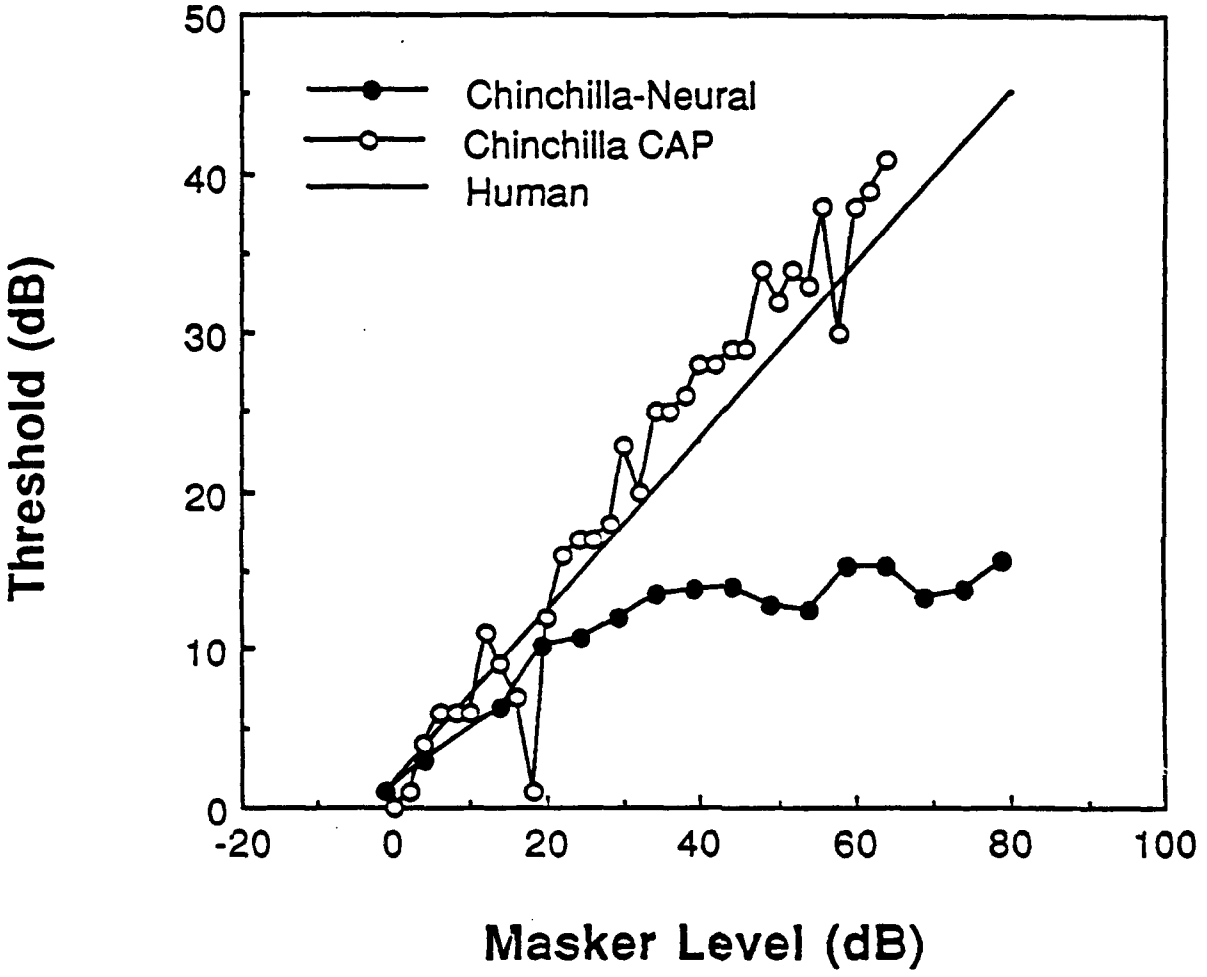

Chinchilla-Neural
Chinchilla CAP
Human
Threshold (dB)
Masker Level (dB)
50
40
30
20
10
0
-20
0
20
40
60
80
100

FIGURE 14

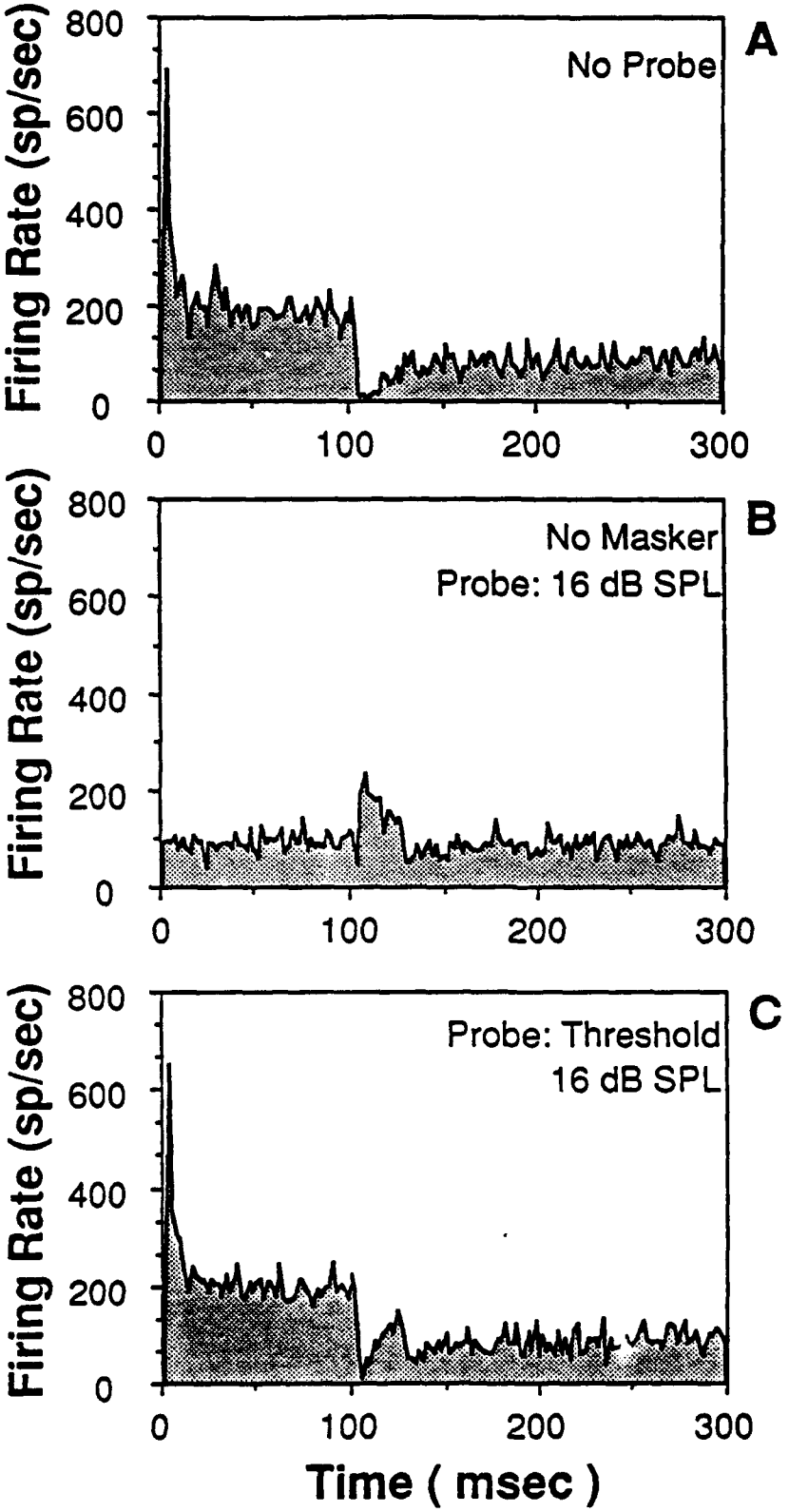

A
No Probe
B
No Masker
Probe: 16 dB SPL
C
Probe: Threshold
16 dB SPL
Firing Rate (sp/sec)
800
600
400
200
0
0
100
200
300
Time (msec)

FIGURE 15

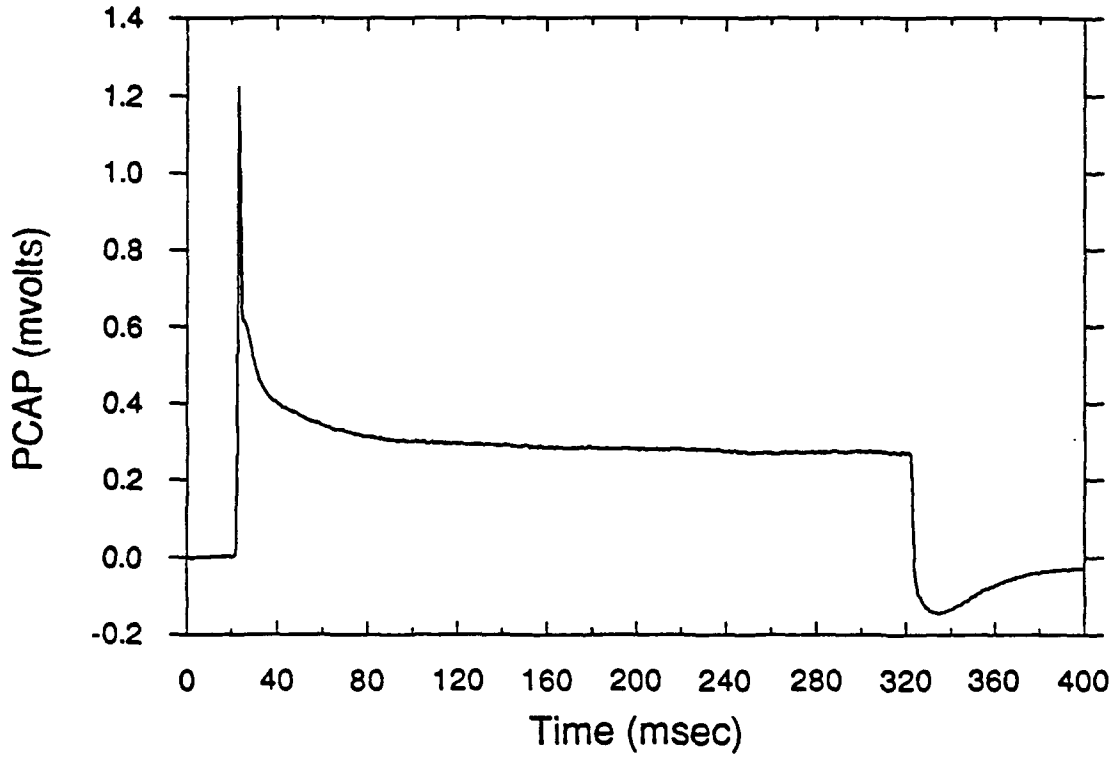

1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
PCAP (mvolts)
0
40
80
120
160
200
240
280
320
360
400
Time (msec)

FIGURE 16

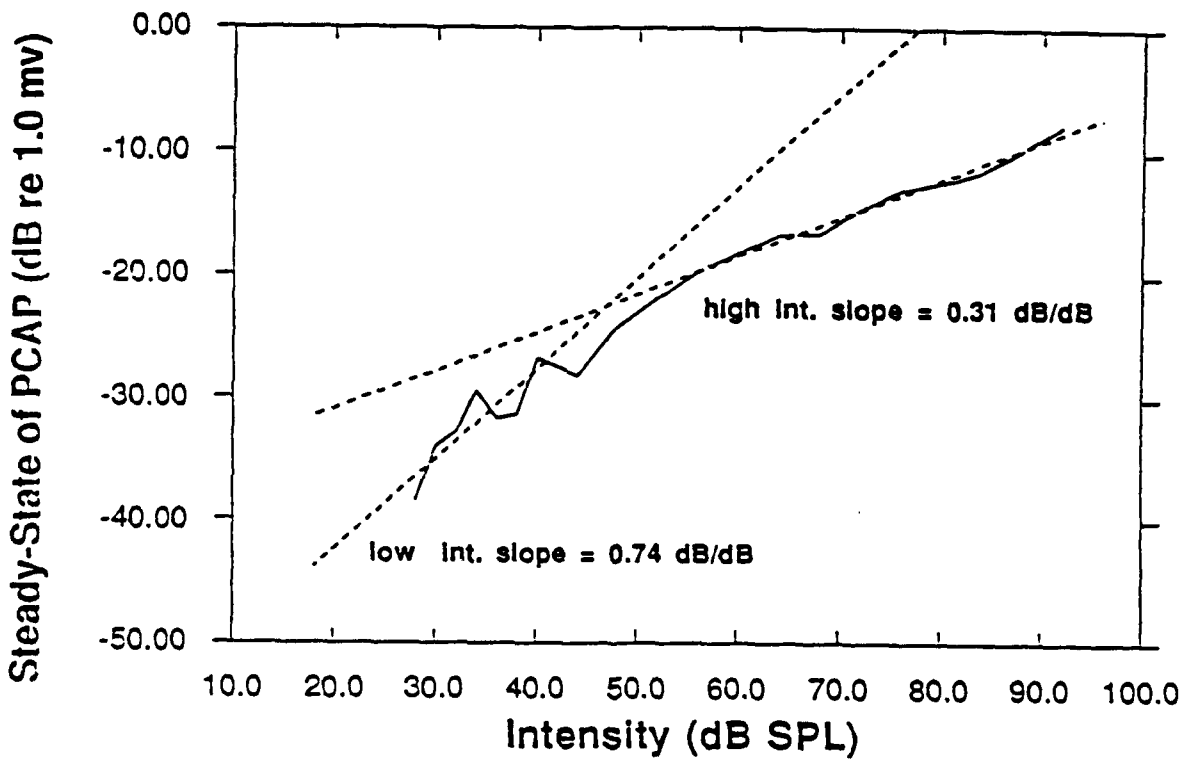

Steady-State of PCAP (dB re 1.0 mv)
0.00
-10.00
-20.00
-30.00
-40.00
-50.00
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
Intensity (dB SPL)
high int. slope = 0.31 dB/dB
low int. slope = 0.74 dB/dB

FIGURE 17

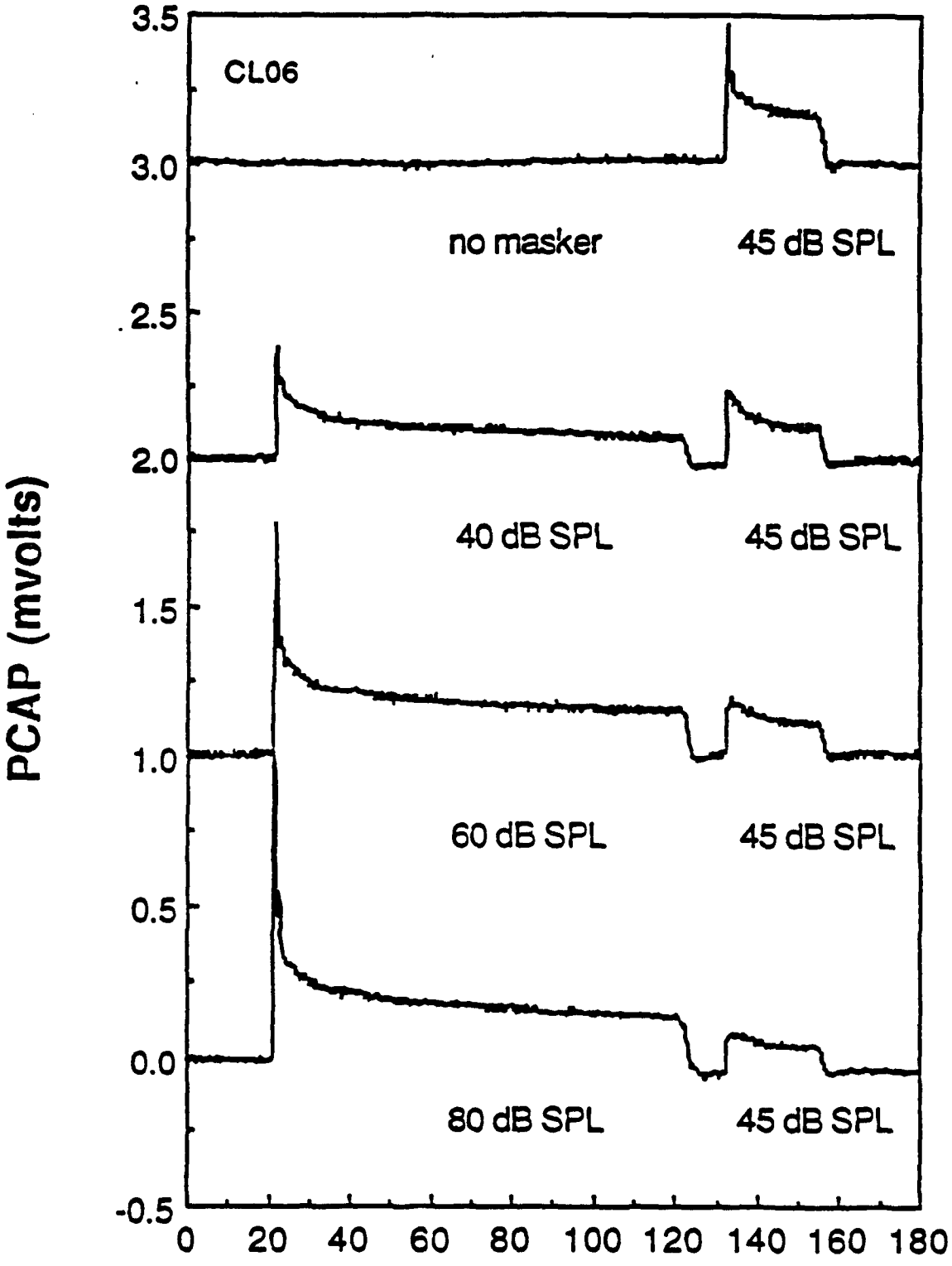

CL06
no masker
45 dB SPL
40 dB SPL
45 dB SPL
60 dB SPL
45 dB SPL
80 dB SPL
45 dB SPL
PCAP (mvolts)
Time (msec)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
0
20
40
60
80
100
120
140
160
180

FIGURE 18

Physiological mechanisms of masking and intensity discrimination

Bertrand Delgutte

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, Department of Otolaryngology, Harvard Medical School, and Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA 02114.

Presented at 117th Meeting of the Acoustical Society of America, Syracuse, NY, May 23 1989.

In this talk, I will report physiological and modeling studies designed to test an hypothesis that has played a key role in hearing theory for the last 50 years. I am referring to Fletcher's idea that masking patterns produced by acoustic stimuli reflect the pattern of neural excitation produced by the masker as a function of place along the cochlea. Two conditions are necessary for this hypothesis to be true: First, masking must be due to the spread of the excitation produced by the masker to the place of the signal along the cochlea. Second, the tone signal must be detected by neurons tuned to its frequency. Both of these assumptions have been questioned: First, masking might be due to the suppression ot the neural response to the signal by the masker, even if the masker does not excite neurons tuned to the signal frequency. Second, psychophysical evidence suggests that listeners do not attend to neurons tuned to the signal frequency if other neurons provide better detectability. This phenomenon is known as "off-frequency listening".

The purpose of the studies I will now describe was to identify the roles of suppression and off-frequency listening in masking. First, I will describe physiological observations on the masking effect of a 1-kHz tone on the responses of auditory-nerve fibers in anesthetized cats. Then I will show masking results for a broader range of stimulus conditions using a model of peripheral auditory processing.

The method I have used for separating the contributions of suppression and spread of excitation to masking originates in Houtgast's suggestion that differences in masked thresholds obtained by simultaneous and nonsimultaneous masking techniques measure the contribution of suppression to masking because suppression only occurs when the signal and the masker are simultaneously presented. The FIRST SLIDE illustrates how suppression and spread of excitation can mask the response of an auditory-nerve fiber. The top panel shows a schematic tuning curve (dashed lines) and boundaries of the two-tone suppression area (thick lines) for an auditory-nerve fiber. Together, these two curves divide the frequency-intensity plane into 3 regions: Region E in which tones can only excite auditory-nerve fibers, Region S in which tones can only suppress, and Region ES in which tones both excite and suppress. The bottom panels show discharge rate as a function of the level of a tone signal, both in the presence and in the absence of fixed maskers placed in each of the three regions. The left panel shows the excitatory masking which occurs when the fixed masker is placed in Region E. The masker by itself excites the fiber (dashed line). There is no suppression because the response to the signal plus the masker (continuous line) is always greater than the response to the signal alone (dotted line). The masked threshold (circle) is the signal level for which the response to the signal plus the masker just exceeds the response to the masker alone. There is masking because the masked threshold is greater than the threshold in quiet (plus). The center panel shows the suppressive masking which occurs when the masker is placed in region S. The masker produces by itself no response, but shifts the rate-level function for the signal towards high intensities, resulting in a threshold increase. Excitatory and suppressive masking are not mutually exclusive, as shown in the right panel. The masker, placed in region ES, excites the fiber and suppresses the signal, as shown by the fact that the response to the signal plus the masker is smaller than the response to the signal alone over a range of levels.

One can discriminate between these three forms of masking by making 3 threshold measurements from auditory nerve fibers. Two of these thresholds have already been defined: the threshold in quiet (plus), and the simultaneous masked threshold (circle). The third type of threshold (triangle), which I

call the *nonsimultaneous threshold,* is the signal level for which the response to the signal alone just exceeds the response to the masker alone. The nonsimultaneous threshold is similar to the pulsation threshold in psychophysics, and to the "isorate probe level" that Young and Hellstrom have recently measured in auditory-nerve fibers. Like the pulsation threshold, it is not, strictly speaking a masked threshold because the signal remains detectable below threshold. Its significance is that *the difference between simultaneous and nonsimultaneous thresholds measures the contribution of suppression to masking* because two-tone suppression only occurs for stimuli that overlap in time. It is clear from the figure that the relative positions of the three thresholds (quiet, simultaneous, and nonsimultaneous) uniquely determine which of the three forms of masking occurs.

The NEXT SLIDE shows how masked thresholds were measured for auditory-nerve fibers in anesthetized cats. The top panel shows the method for measuring simultaneous thresholds, which mimics the two-interval, two-alternative forced choice paradigm of psychophysics. Two stimuli are presented repeatedly in random order. One consists a tone signal plus a masker, while the other one is the masker alone. The number of spikes produced by each stimulus is counted, and the signal level adjusted by a PEST procedure so that the spike count for the signal plus the masker exceeds the count for the masker alone in 75% of the stimulus presentations. The bottom panel shows the method for measuring nonsimultaneous thresholds. It is basically the same as that used for simultaneous thresholds, except that the signal *alone* alternates with the masker.

The NEXT SLIDE summarizes masked-threshold measurements from many auditory-nerve fibers for a 1-kHz tone masker at 60 dB SPL. Masked thresholds were measured for dozens of auditory-nerve fibers with different CF's for each signal frequency in a fixed set. For any given masker and signal frequency, it seems likely that the fibers which provide most information for detecting the signal are those that have the lowest masked thresholds. Thus, each symbol in the top panel represents the masked threshold of the auditory nerve fiber that had the lowest threshold for the signal frequency corresponding to the abscissa, regardless of its CF. The resulting *best threshold patterns* resemble psychophysical masking patterns for a 60-dB tone masker: Both show a maximum at the masker frequency, and a skew towards high frequencies, meaning that the 1-kHz masker is more effective in masking high-frequency signals than low-frequency signals. For signal frequencies above the masker, simultaneous masked thresholds (circles) are higher than nonsimultaneous thresholds (triangles), indicating that suppression contributes to masking. In contrast, for signal frequencies near and below the masker, masked thresholds are similar for the two conditions, indicating that masking is due to the spread of the excitation produced by the masker.

The bottom panel addresses effects of off-frequency listening. If there were no off-frequency listening, the CF of the fiber that has the lowest masked threshold would always coincide with the signal frequency. Here, the difference between the CF of the best fiber and the signal frequency is plotted as a function of signal frequency. This difference is expressed in units of percent distance along the cochlea by means of Liberman's cochlear frequency map. A positive difference means that the best CF is higher in frequency than the signal. For both simultaneous and nonsimultaneous conditions, the frequency offset is positive for signal frequencies above the masker, and negative for signal frequencies below the masker. The direction of this offset is predictable from a linear filter bank model of signal detection, and is consistent with psychophysical estimates of off-frequency listening.

The NEXT SLIDE shows improvements in signal detectability provided by off-frequency listening. Circles refer to the best threshold pattern for the 60-dB, 1-kHz, nonsimultaneous masker reproduced from the previous slide. Triangles show the masked thresholds of the auditory-nerve fibers whose CF's coincide with the signal frequency. These *on-frequency* thresholds are about 10-15 dB higher than the best thresholds for signal frequencies near the masker. Thus, the ability to listen off-frequency significantly improves signal detectability when the signal is close in frequency to the masker. For signal frequencies far from the masker, the role of off-frequency listening is minimal. On-frequency threshold patterns realize a suggestion by Moore for characterizing the response of the

auditory nerve in a manner consistent with the psychophysical concept of excitation.

The NEXT SLIDE shows best threshold patterns in simultaneous and nonsimultaneous masking for the 1-kHz masker at three different levels. The center panel reproduces results for the 60-dB masker from a previous slide. For the 40 dB masker (left panel), masking is restricted to a narrower range of signal frequencies than at 60 dB. Thresholds are similar for both masking conditions, indicating that suppression plays a minimal role at this level. When the masker level is raised to 80 dB (right panel), simultaneous masking extends much farther towards high frequencies than at the lower levels. This rapid growth of masking resembles the upward spread of masking in psychophysics. A key question is whether upward spread of masking is due to an increase in suppression or in spread of excitation. The difference between simultaneous and nonsimultaneous thresholds, which measures suppression, is much greater at 80 dB than at 60 dB, showing that suppression grows rapidly. In contrast, nonsimultaneous masking patterns, which do not include effects of suppression, grow only moderately for the 20-dB increase in masker level. Thus, I conclude that upward spread of masking is largely due to the rapid growth of suppression rather than to an increase in the excitation produced by the masker.

In summary, this physiological experiment has shown that, for 1-kHz tone maskers, patterns of best auditory-nerve fiber masked threshold against signal frequency resemble psychophysical masking patterns in many respects. Simultaneous masking is due to both suppression of the signal by the masker and spread of the excitation produced by the masker, with relative contributions of two masking mechanisms depending on the frequency separation between the signal and the masker. Suppression dominates for signal frequencies well above an intense masker, and in particular is largely responsible for the upward spread of simultaneous masking. Off-frequency listening considerably contributes to signal detection for signal frequencies near the masker.

In the remainder of this talk, I will describe results for a model of the peripheral auditory system designed to simulate masking experiments for a broader range of stimulus conditions than would be practical to study in physiological experiments. The model is an elaboration of one I presented 3 years ago to predict intensity discrimination. The model consist of approximately 100 hundred frequency channels whose center frequencies range from 100 Hz to 50 kHz in steps of 1/12 octave. The NEXT SLIDE shows a block diagram for one model channel. Because the model works in the frequency domain, it takes as input the power spectrum the sound stimulus. The input spectrum is processed by two filters, an excitation filter and a suppression filter. The bottom left panel shows that the excitation filter (thin line) resembles an inverted tuning curve, while the suppression filter (thick line) resembles an inverted suppression threshold curve. The power at the output of the suppression filter is the input to a nonlinear suppression growth function function which outputs a variable which I call "suppression potential". This suppression potential may or may not result in actual suppression of the final model response depending on the stimulus spectrum. The suppression growth functions, which are shown in the bottom panel, are the same that were used by Sachs and Abbas in their model of suppression. The rate of growth of suppression depends on the position of the center of gravity of the signal spectrum with respect to the CF of the channel: If the center of gravity is below CF (thick line), suppression grows faster than 1 dB/dB, while the rate is less than 1 dB/dB for stimuli near and above the CF (thin line). The resulting suppression potential modifies the tuning of the excitation filter. The bottom right panel shows the excitation filter in the absence of suppression, and with suppression potentials of 10, 20, 30 and 40 dB. The effect of suppression is greatest for frequencies near the CF, consistent with physiological observations of Kiang & Moxon, Schmiedt, and Fahey & Allen. The power at the output of the excitation filter is the input to a nonlinear function describing the growth of discharge rate with intensity. This growth function produces two outputs: the mean discharge rate, and the variance of the discharge rate over repeated presentation of the stimulus. The variance is essential for measuring discriminability in simulations of masking experiments.

While the excitation and suppression filters are the same for all the model fibers that constitute one frequency channel, the rate growth functions differ for each fiber in order to simulate the

variability in thresholds, SR and dynamic ranges that has been observed by Liberman in auditory-nerve fibers. The NEXT SLIDE shows the rate-level functions for the 9 model fibers which constitute one frequency channel. These growth functions are based on fits to rate-level functions from many auditory-nerve fibers, using a function proposed by Sachs, Abbas, and Winslow. The bottom panel shows how the rate variance varies as a function of the mean rate. This curve is based on fits to physiological data based on a Poisson model with deadtime.

The NEXT SLIDE shows results for a 2-kHz fiber of the model. The thin line in the top panel shows a threshold tuning curve measured by a tracking procedure similar to that used in physiological experiments. Thick lines shows boundaries of the suppression area measured using a tracking procedure similar to that of Schmiedt. The model's suppression areas resemble those observed by Sachs and Schmiedt in auditory-nerve fibers. The bottom panel shows rate-level functions for 3 tones of different frequencies in the presence of a fixed tone at the CF. When the swept tone is at the CF (triangles), discharge rate increases over that produced by the fixed tone, meaning that the is no suppression. For tones whose frequencies are either below (circles) or above (plusses) the CF, discharge rate first decreases as the swept tone enters the suppression area, then increases when the swept tone reaches the tuning curve threshold. This behavior closely resembles that found in auditory-nerve fibers by Sachs and Abbas.

The NEXT SLIDE compares physiological masking data with model predictions for a 1-kHz tone masker at 60 dB SPL. The right panel reproduces the physiological best threshold patterns from a previous slide. The left panel shows best threshold patterns for the model. In simultaneous masking, each model best threshold was obtained by computing the responses of all 900 model fibers to both the signal plus the masker and the masker alone, then adjusting the signal level so that the discriminability index d' between the responses to the two stimuli is exactly 1 for the best fiber. This corresponds to the same 75% correct detection criterion for a 2I-2AFC procedure that was used in the physiological experiments. Nonsimultaneous best thresholds are obtained by a similar method, except that the model's response to the signal alone is compared with the response to the masker.

The general shapes of best threshold patterns are similar for the model and the physiological data. In both cases, thresholds are greater in simultaneous than in nonsimultaneous masking for signal frequencies above the masker, indicating a contribution of suppression to masking. An obvious difference is that suppression occurs for signal frequencies below the masker in the model, but not in the physiological data. This is how I explain this discrepancy: Suppression for low-CF fibers shows great variability between cats. In measuring the suppression threshold curves that were used in constructing the model, I necessarily selected those fibers that showed the strongest suppression. No such selection was applied in measuring masked thresholds shown at the right. Thus, the model is likely to be representative of cats in which suppression in low-CF fibers is most prominent, and may somewhat overestimate suppression in more typical cats. The bottom panels show that patterns of place offsets, which measure off-frequency listening, are qualitatively similar for the model and the physiological data.

The NEXT SLIDE shows model best threshold patterns for a 1-kHz tone at three different levels. The center panel reproduces the 60-dB data from the previous slide. For the 40 dB masker (left), the patterns are more restricted than at 60 dB, and there is no evidence for suppression as for the physiological data. For the 80 dB masker (right), best threshold pattern extends much farther towards high frequencies than at lower levels. The difference between simultaneous and nonsimultaneous thresholds is also greatly increased, consistent with our suggestion that the rapid growth of simultaneous masking is largely due to an increase in suppression. These model results resemble the physiological data, although the difference between simultaneous and nonsimultaneous thresholds is greater in the physiological data than in the model for signal frequencies between 2 and 10 kHz.

Being now assured that the model qualitatively simulates the major effects of masking for 1-kHz tone maskers, we can examine other stimulus situations. In the 15 years since Zwicker gave his well-known paper, the most popular way to measure masking has been psychophysical tuning curves, in

which the signal rather than the masker is held fixed. The NEXT SLIDE compares model psychophysical tuning curves obtained in simultaneous and nonsimultaneous masking with human psychophysical data of Moore, Glasberg & Roberts. For both sets of data, psychophysical tuning curves are shown for three different signal frequencies. Comparison between psychophysical and model data is made difficult by the fact that psychophysical data are only available for a relatively narrow range of frequencies. Furthermore, human psychophysical tuning curves are sharper than model tuning curves, consistent with previous evidence that frequency selectivity is greater in humans than in cats. If we ignore these species differences and focus on the differences between the two masking conditions, it is clear that, for both model and human data, psychophysical tuning curves are sharper in nonsimultaneous masking (thin lines) than in simultaneous masking (thick lines), with the largest differences occurring in the skirts rather than in the tips of the tuning curves.

A question of great interest to psychophysicists is whether psychophysical tuning curves accurately reflect the tuning characteristics of single auditory-nerve fibers. The NEXT SLIDE compares nonsimultaneous model psychophysical tuning curves (which represent the best thresholds among 900 model fibers) with the threshold tuning curves of single model fibers whose sensitivity were adjusted so that they would coincide with the PTC thresholds at their tips. The two types of tuning curves closely superimpose, with the exception of the tails of high-frequency tuning curves for which psychophysical tuning curves (thin lines) are slightly higher. This close correspondence between psychophysical and single-fiber tuning curves is due to the fact that off-frequency listening is minimal for the low-level signals used in these threshold measurements.

A psychophysical correlate of suppression originally found by Houtgast is that the masking produced by a two-component masker can be smaller than that produced by one of its two components. Such *unmasking* effects are found in nonsimultaneous masking, but not in simultaneous masking. The bottom panel of the NEXT SLIDE shows data from one of Houtgast's experiments. The horizontal line shows the masked threshold for a 1-kHz tone signal when the masker is a 1-kHz tone at 40 dB SPL. Symbols show the threshold for the same signal when a variable-frequency tone at 60 dB SPL is added to this fixed 1-kHz masker. When the signal is presented simultaneously with the two-tone masker (circles), masked thresholds are always greater for the two-tone masker than for the fixed-tone masker. In contrast, when the signal and the masker are nonsimultaneous (triangles), there is a range of frequencies both above and below 1 kHz in which the two-tone masker produces less masking than the fixed tone, indicating that the signal is unmasked. The top panel shows a model simulation of Houtgast's experiment. Stimuli were the same as in the psychophysical experiment, except that the fixed masker and the signal were at 2 kHz rather than 1 kHz. As in the psychophysical data, unmasking is found only when the signal and the two-tone masker are not simultaneous presented, although the frequency range of unmasking is broader for the model than for the human data.

A motivation for the present work was to test Fletcher's hypothesis that masking patterns reflect the excitation pattern produced by the masker. Several investigators have applied this idea to determinate how the spectra of speech sounds might be represented at peripheral stages of the auditory system. The NEXT SLIDE compares simultaneous and nonsimultaneous masking patterns produced by the vowel [ae] for the model and for psychophysical data of Moore and Glasberg. For both masking conditions, the peaks in the masking patterns at the formant frequencies are sharper in the human data than for the model, consistent -again- with the notion that humans ears are more frequency selective than feline ears. Nevertheless, in both human and model data, the formant peaks are sharper in nonsimultaneous masking than in simultaneous masking.

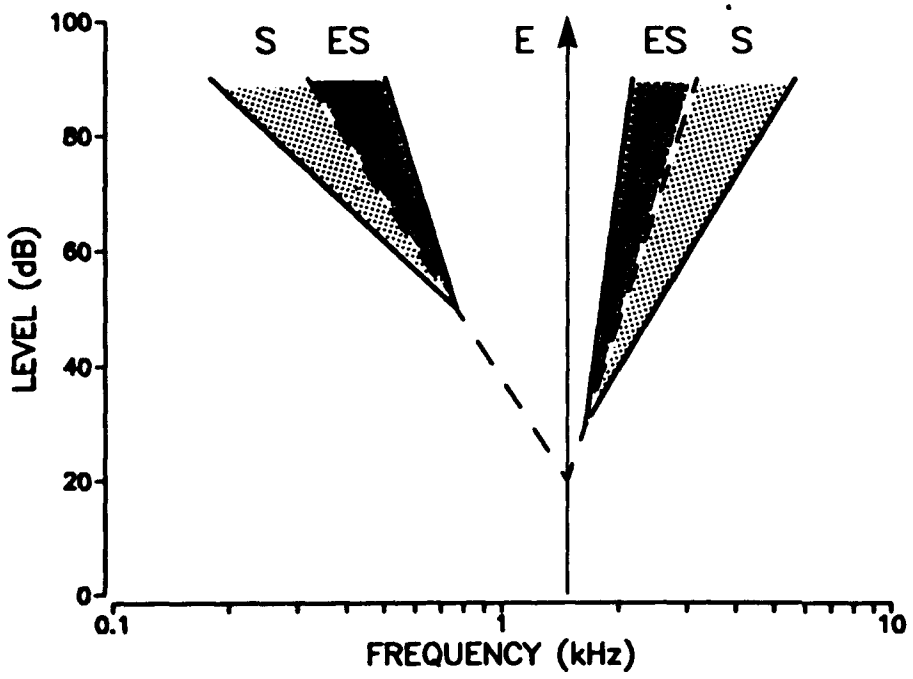
In conclusion, both physiological experiments and model simulations show that suppression plays an important role in masking. Some of the effects of suppression are summarized in the NEXT SLIDE. In order to understand these effects, it helps to distinguish suppression among frequency components of a complex masker ("within-masker suppression") from the suppression exerted by the masker on the signal ("masker-to-signal suppression"). This distinction does not mean that there exist

different mechanisms for suppression, but that the functional consequences of suppression depend on the nature of the psychophysical task. On the basis of experiments with two-tone maskers, we have seen that within-masker suppression produces unmasking in the nonsimultaneous condition, but not in the simultaneous condition. This result fits with the widely accepted view that effects of suppression are only revealed in nonsimultaneous masking, provided that this statement applies to within-masker suppression. Masker-to-signal suppression does not occur in nonsimultaneous masking. In simultaneous masking, however, it causes a broadening of the skirts of the masking patterns, and is largely responsible for the upward spread of masking. Under certain circumstances which we did not have time to describe, a signal can suppress the response of auditory-nerve fibers to a simultaneously-presented masker. Model simulations suggest that such "signal-to-masker" suppression can improve signal detectability in the presence of broadband maskers.

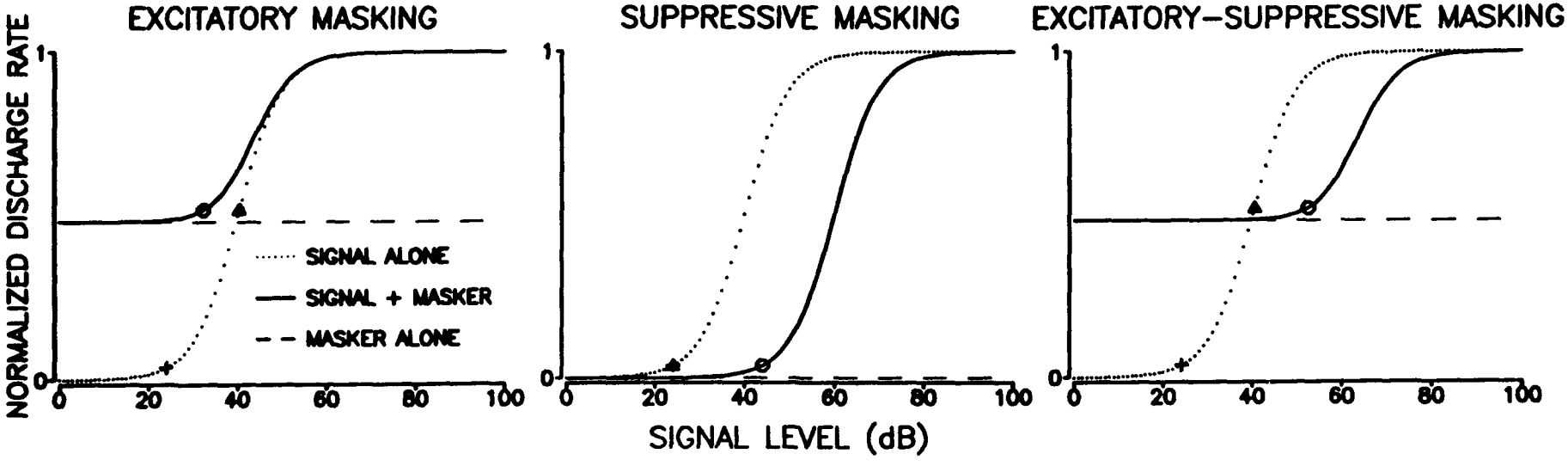
An other factor which has been studied in this talk is off-frequency listening. The physiological experiments showed that off-frequency listening significantly contributes to signal detection for moderate to intense tone maskers. On the other hand, model simulations suggest that off-frequency listening is minimal for the low signal levels used for measuring psychophysical tuning curves, and for broadband noise maskers (not shown). Thus, off-frequency listening seems to improve signal detection for intense, band-limited maskers.

At the beginning of this talk, I argued that Fletcher's hypothesis that masking patterns reflect the pattern of neural excitation produced by the masker along the cochlea can only be true if two conditions are verified: that masking is due to spread of excitation and that off-frequency listening is minimal. Putting together our results on the roles of suppression and off-frequency listening in masking, I conclude that masking patterns can only reveal excitation patterns for nonsimultaneous, broadband maskers.

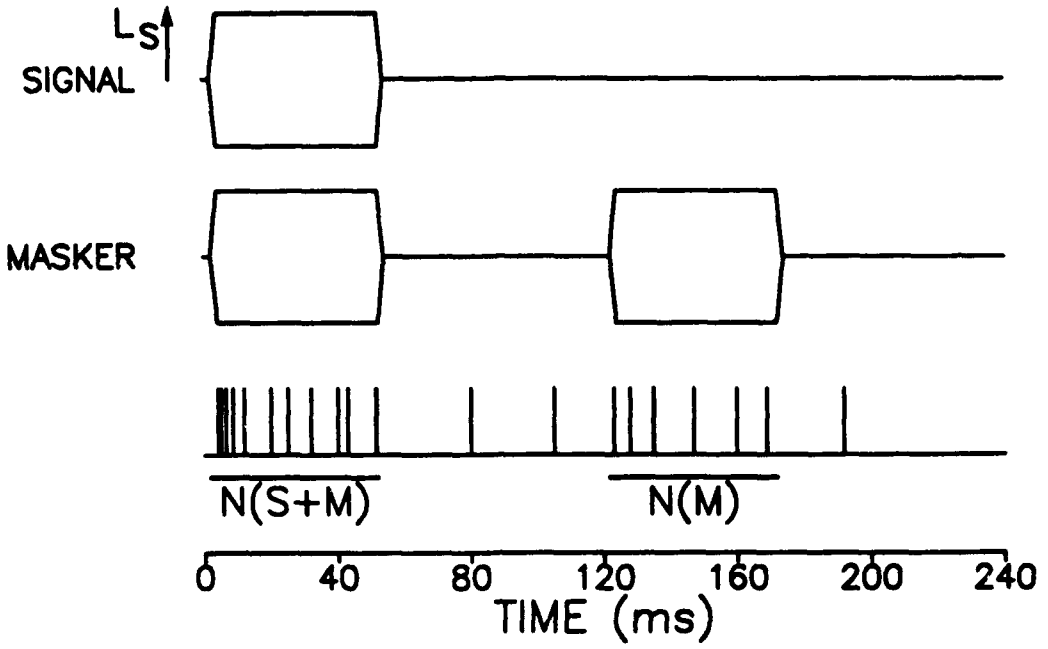
Finally, the overall result of this study is that there is excellent qualitative agreement between psychophysical, physiological, and model measures of masking when important factors such as suppression and off-frequency listening are taken into account. It is remarkable that these results were entirely based on the average discharge rates of auditory-nerve fibers. Thus, most masking phenomena can be simply explained in terms of physiological mechanisms if it is assumed that signal detection is based on average-rate information.



+ THRESHOLD IN QUIET
o SIMULTANEOUS MASKED THRESHOLD
▲ NONSIMULTANEOUS THRESHOLD

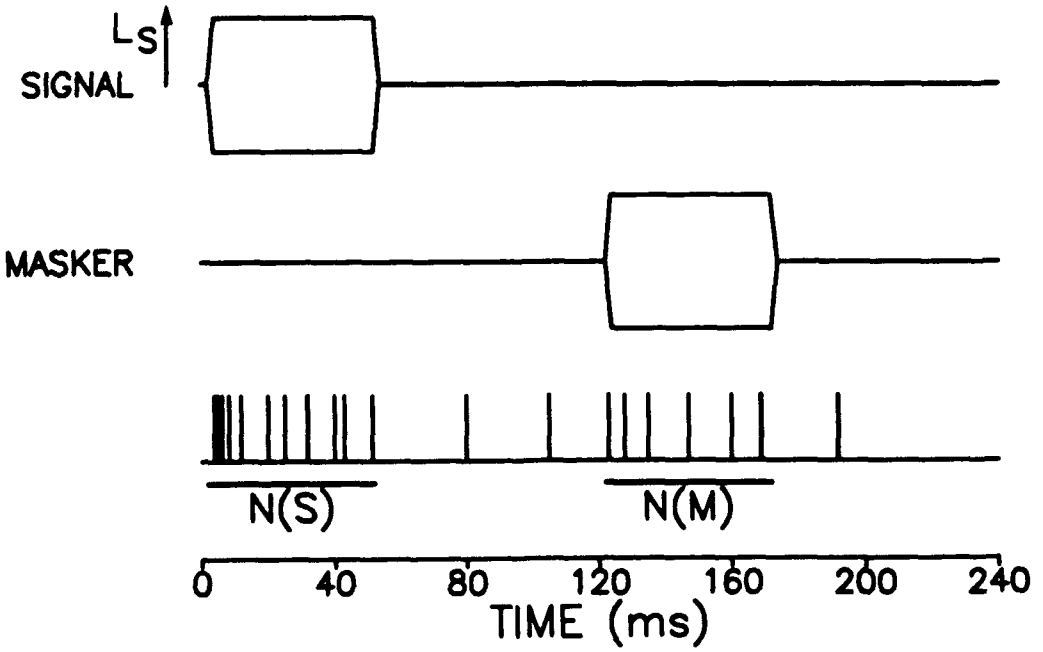


SIMULTANEOUS MASKED THRESHOLD

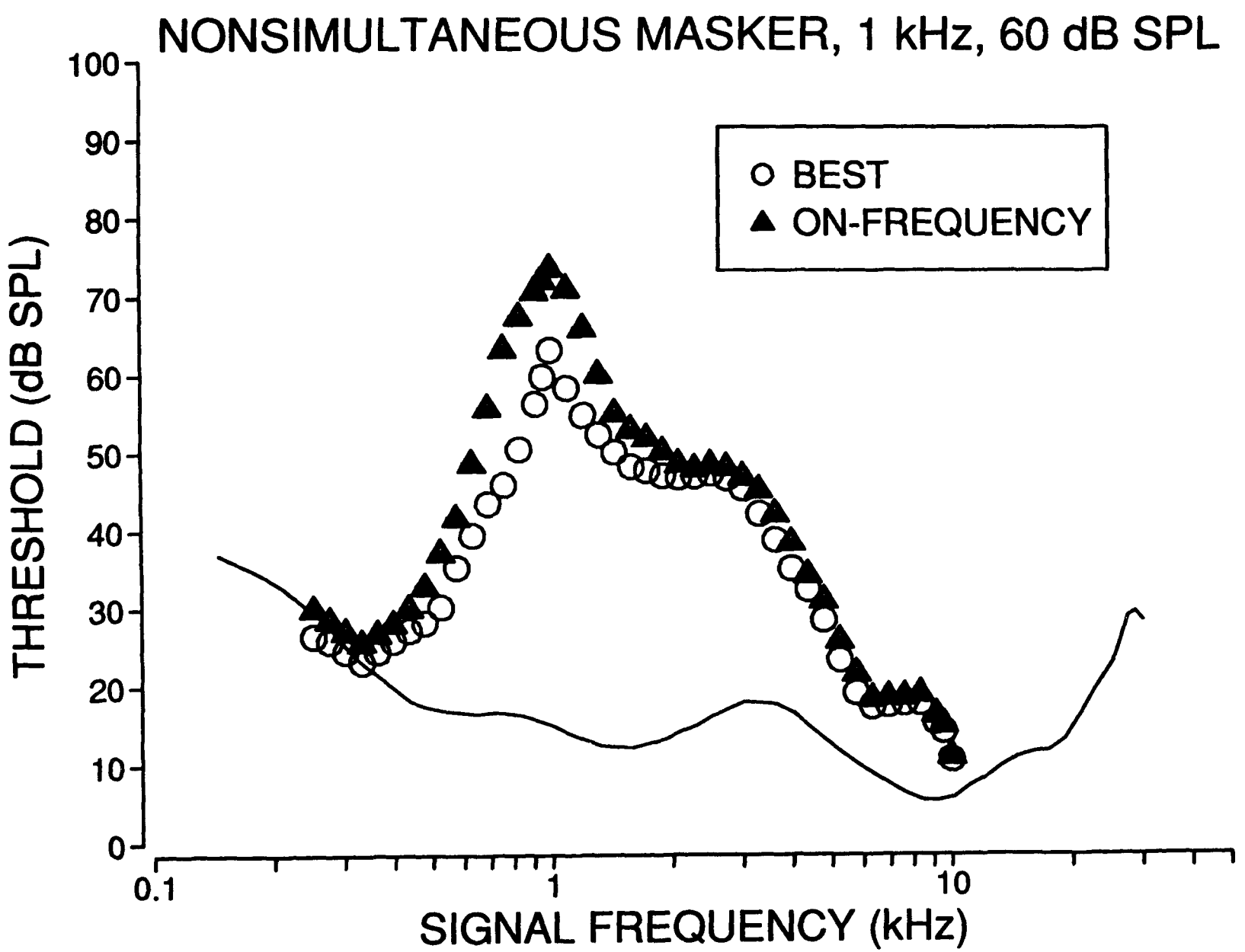


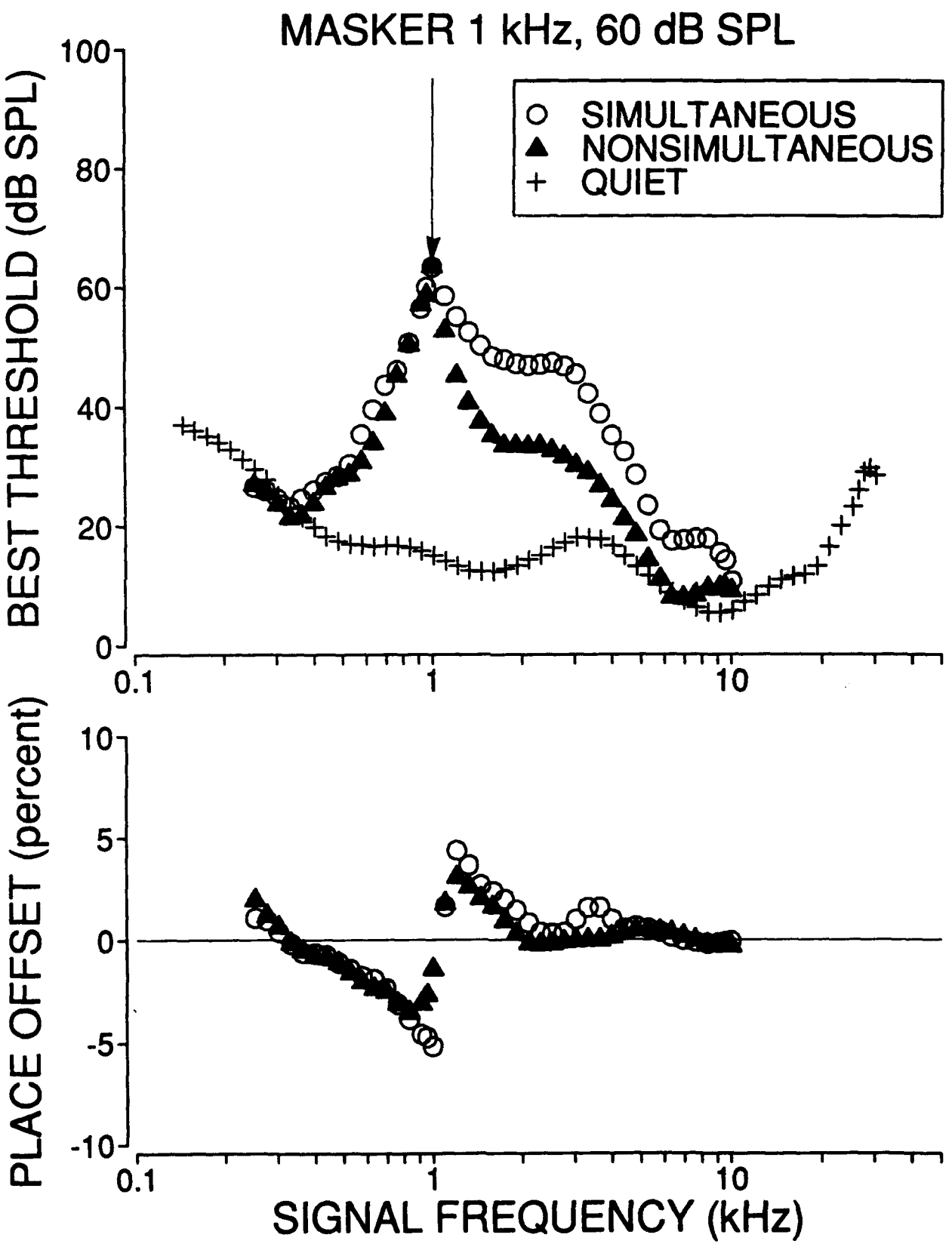
Adjust L_S until Prob[N(S+M) >N(M)] = P_{CR}

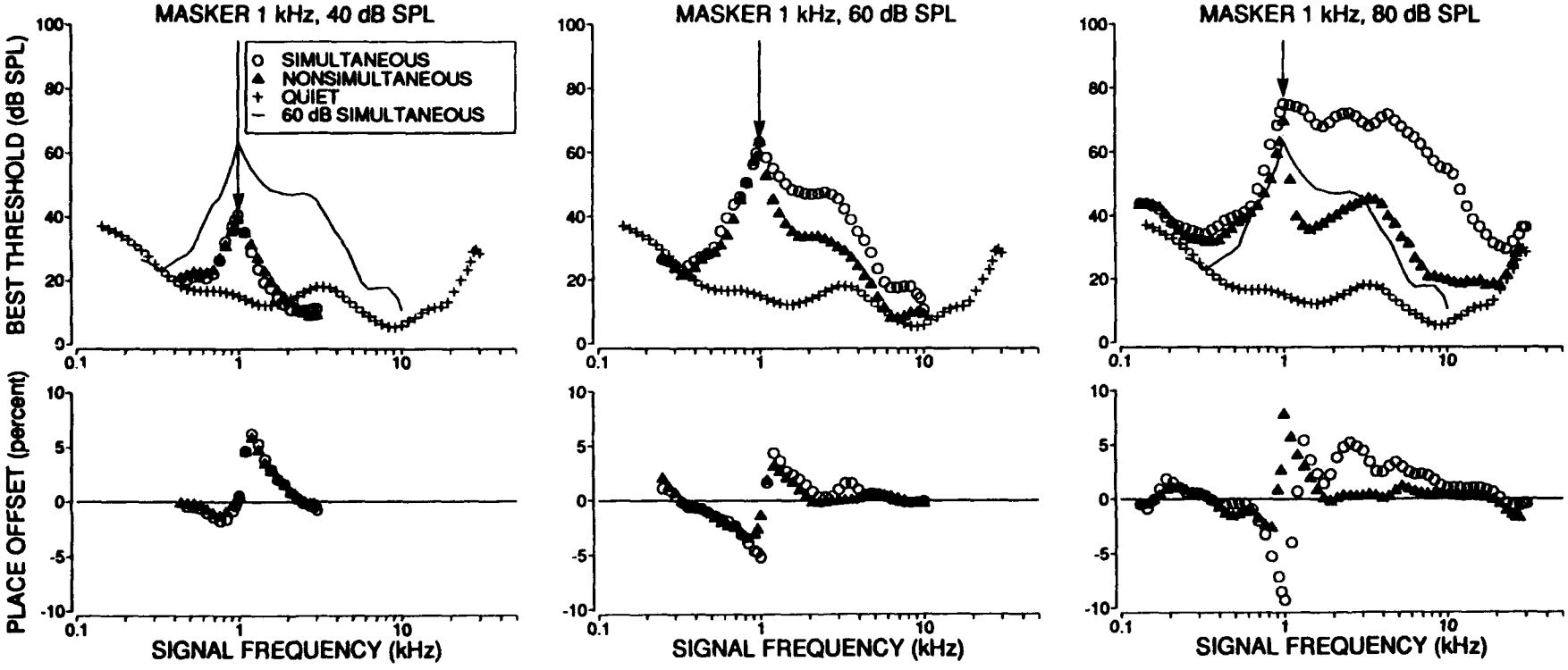
NONSIMULTANEOUS THRESHOLD

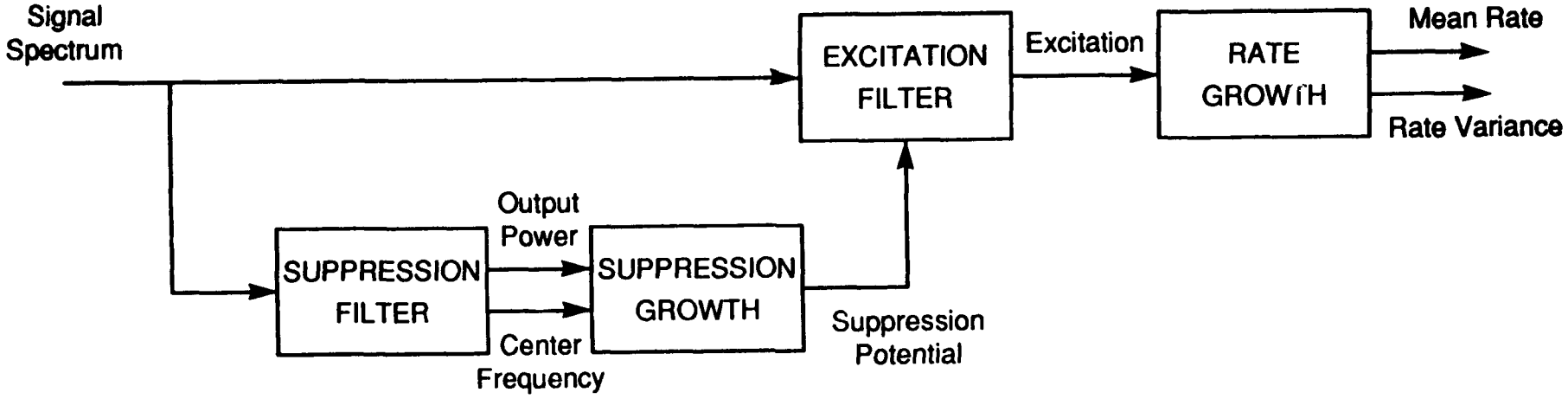


Adjust L_S until Prob[N(S) >N(M)] = P_{CR}

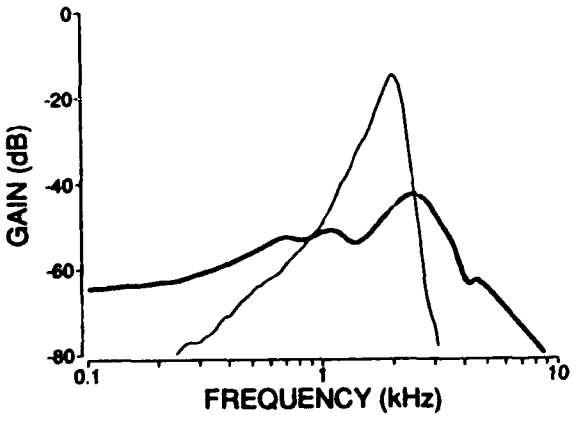

NONSIMULTANEOUS MASKER, 1 kHz, 60 dB SPL
BEST
ON-FREQUENCY
THRESHOLD (dB SPL)
100
90
80
70
60
50
40
30
20
10
0
0.1
1
10
SIGNAL FREQUENCY (kHz)


MASKER 1 kHz, 60 dB SPL
SIMULTANEOUS
NONSIMULTANEOUS
QUIET
BEST THRESHOLD (dB SPL)
100
80
60
40
20
0
0.1
1
10
PLACE OFFSET (percent)
10
5
0
-5
-10
0.1
1
10
SIGNAL FREQUENCY (kHz)

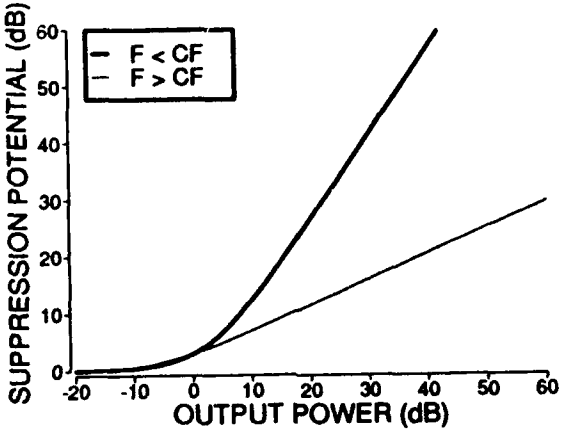

MASKER 1 kHz, 40 dB SPL
MASKER 1 kHz, 60 dB SPL
MASKER 1 kHz, 80 dB SPL
SIMULTANEOUS
NONSIMULTANEOUS
QUIET
60 dB SIMULTANEOUS
BEST THRESHOLD (dB SPL)
PLACE OFFSET (percent)
SIGNAL FREQUENCY (kHz)
SIGNAL FREQUENCY (kHz)
SIGNAL FREQUENCY (kHz)



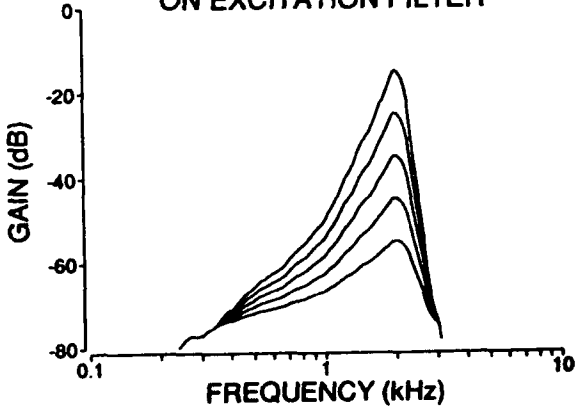
SUPPRESSION AND EXCITATION FILTERS

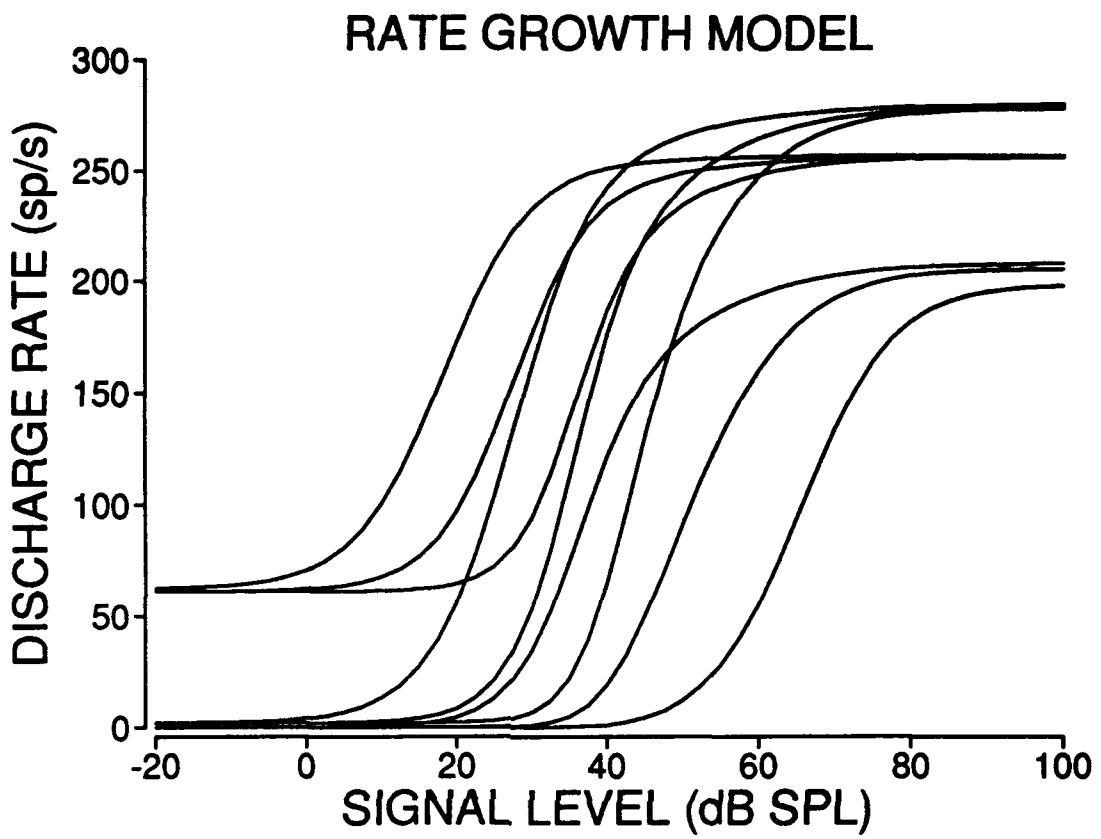


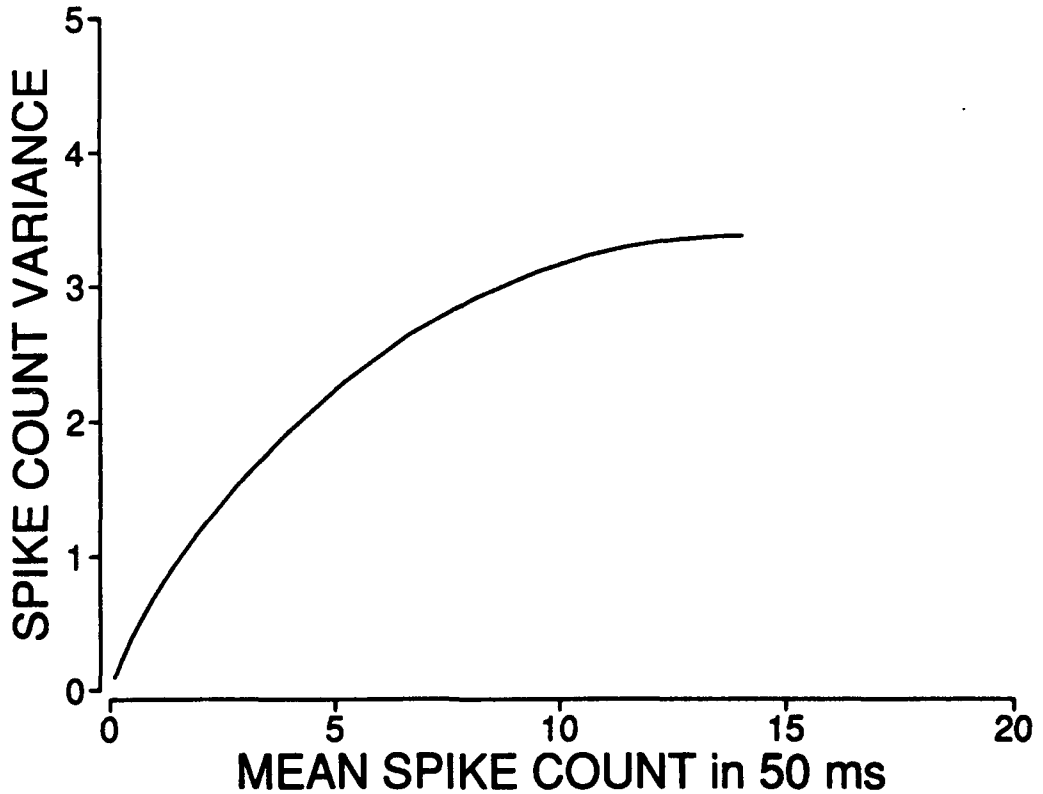
SUPPRESSION GROWTH

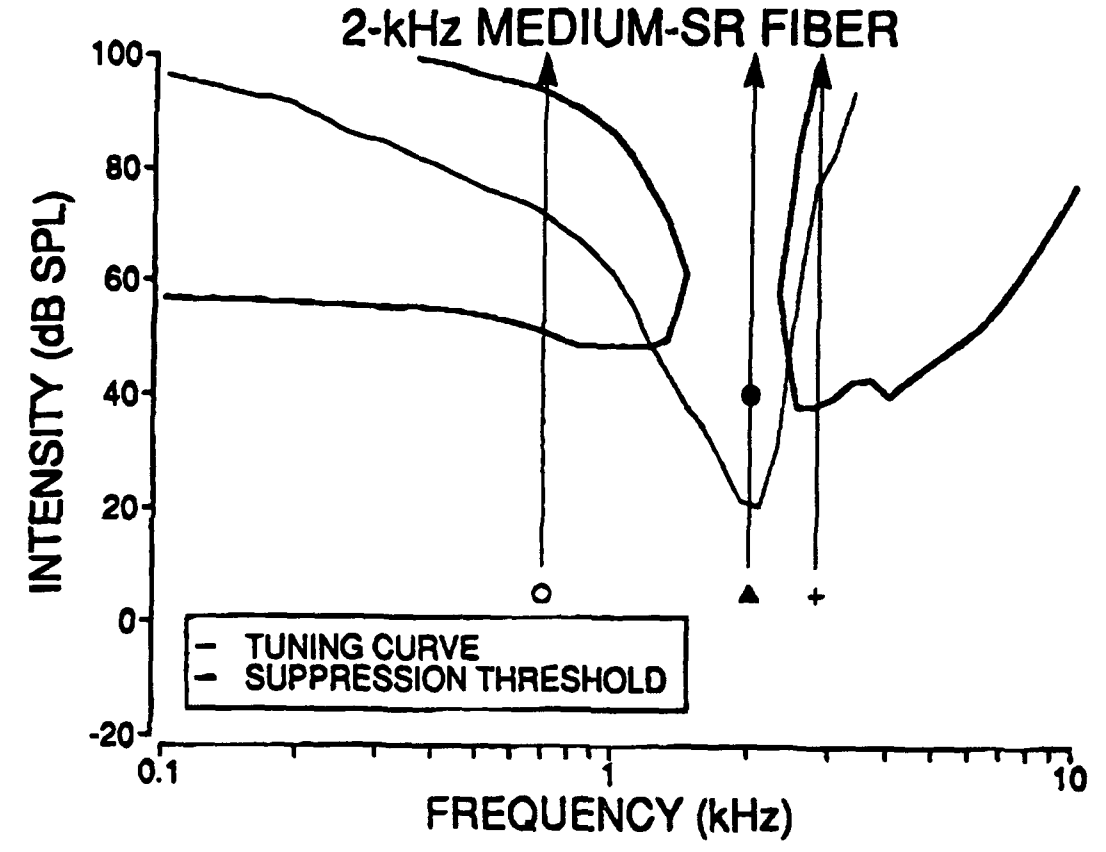


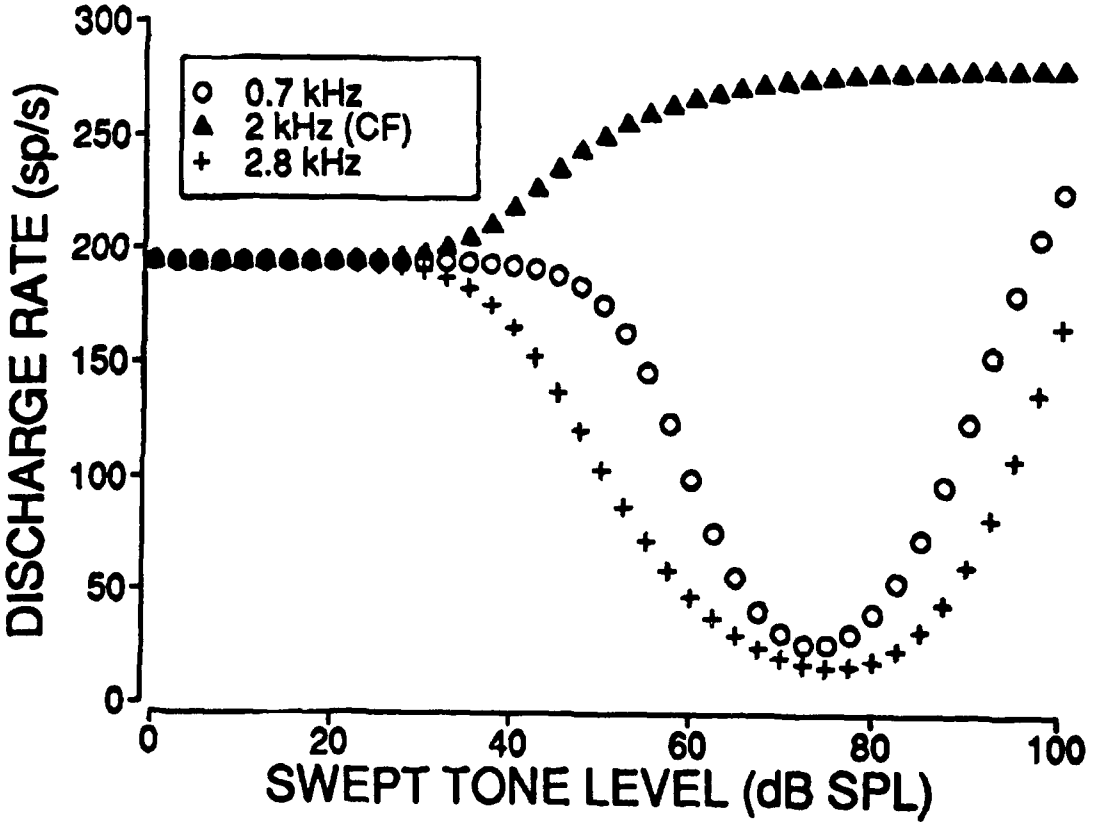
EFFECT OF SUPPRESSION
ON EXCITATION FILTER



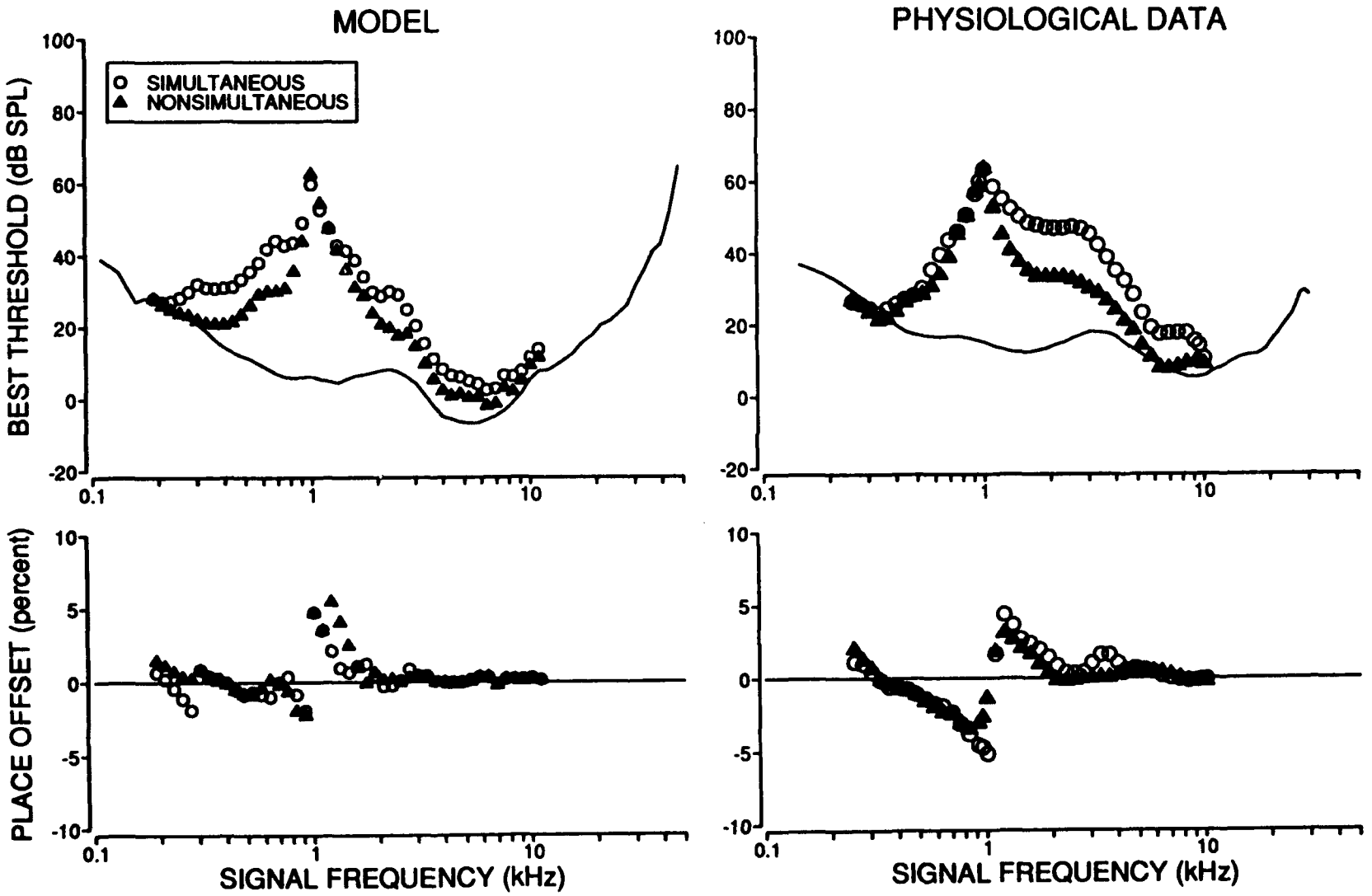

RATE GROWTH MODEL
DISCHARGE RATE (sp/s)
300
250
200
150
100
50
0
-20
0
20
40
60
80
100
SIGNAL LEVEL (dB SPL)

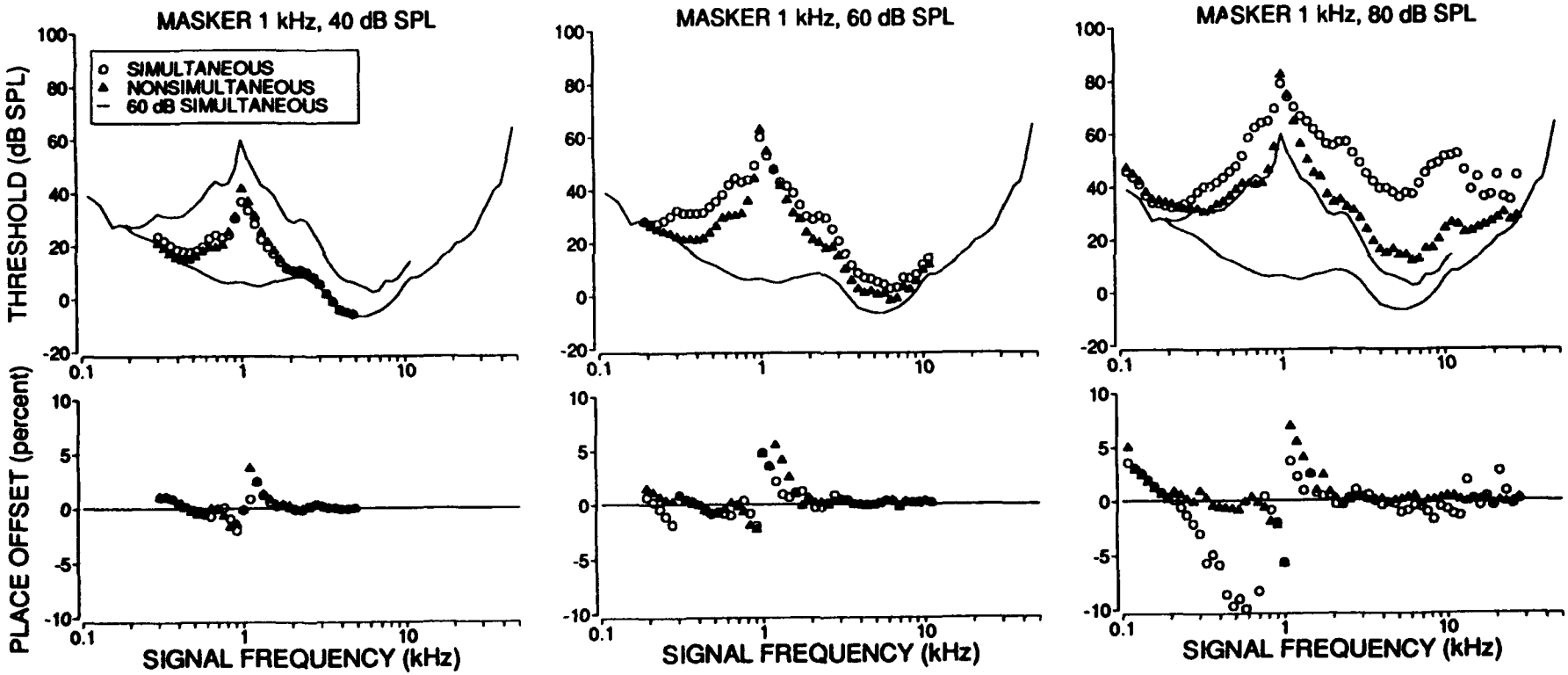

SPIKE COUNT VARIANCE
5
4
3
2
1
0
0
5
10
15
20
MEAN SPIKE COUNT in 50 ms


2-kHz MEDIUM-SR FIBER
INTENSITY (dB SPL)
100
80
60
40
20
0
-20
TUNING CURVE
SUPPRESSION THRESHOLD
0.1
1
10
FREQUENCY (kHz)

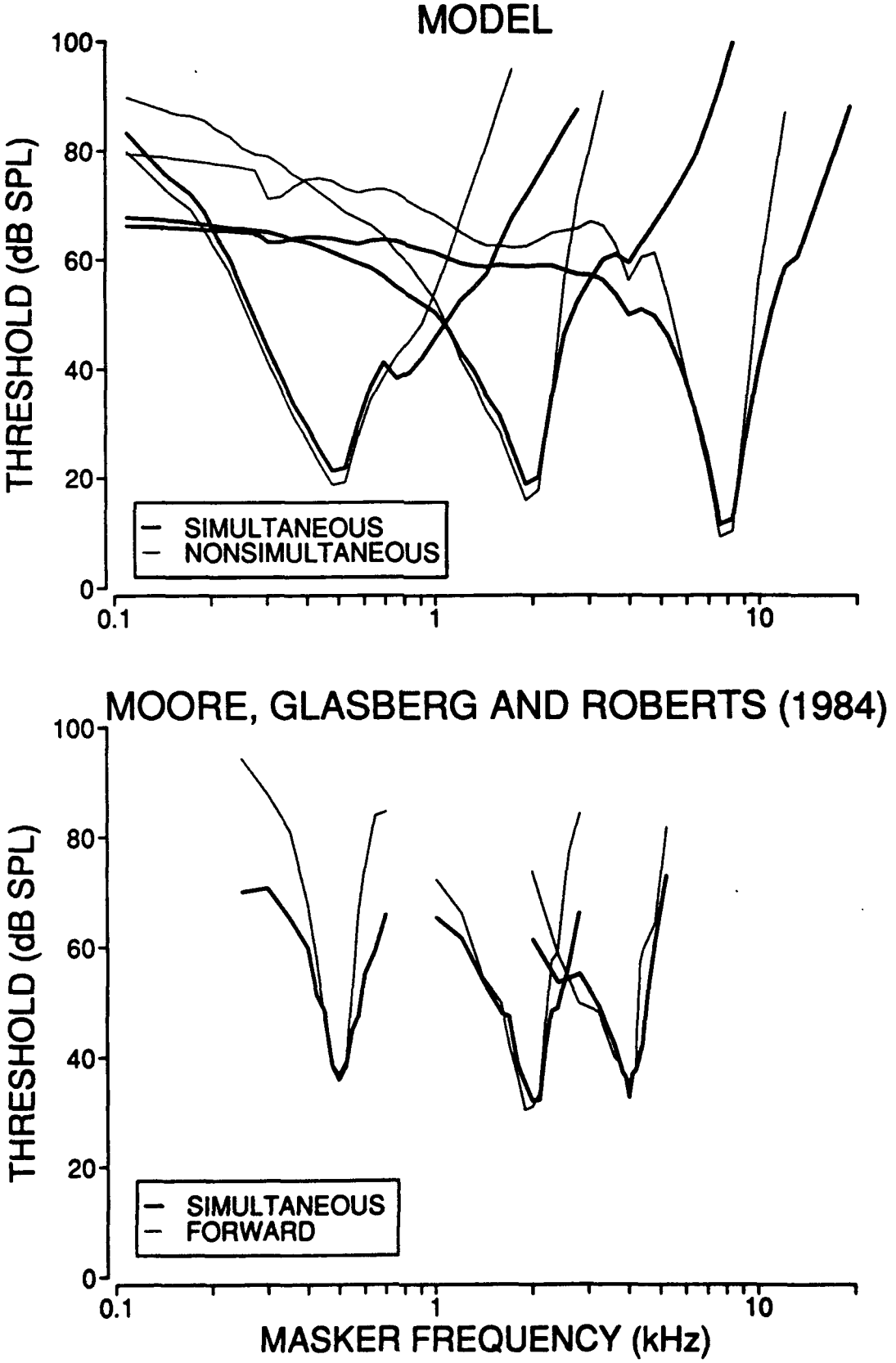

DISCHARGE RATE (sp/s)
300
250
200
150
100
50
0
0.7 kHz
2 kHz (CF)
2.8 kHz
0
20
40
60
80
100
SWEPT TONE LEVEL (dB SPL)

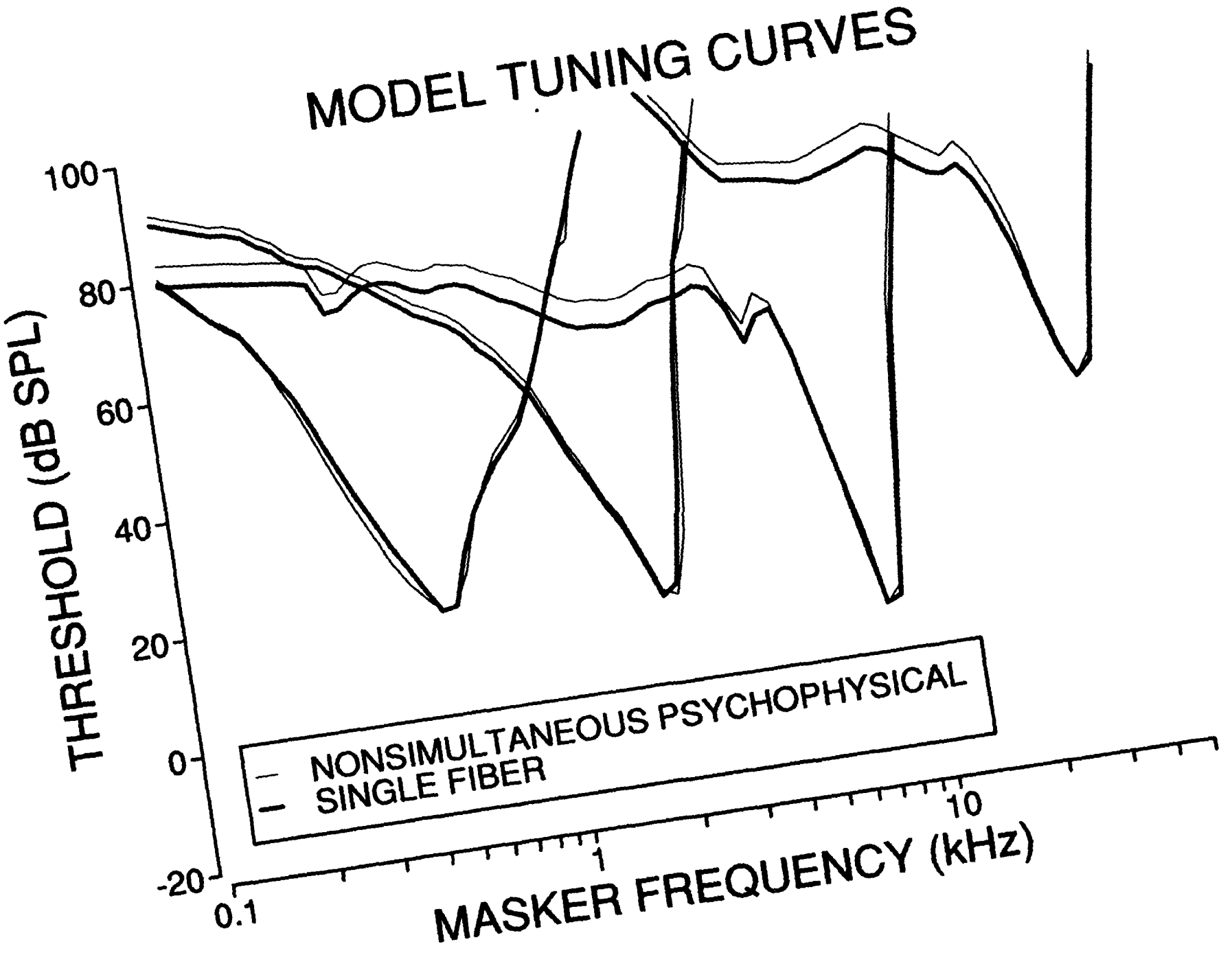
MASKER 1 kHz, 60 dB SPL




MASKER 1 kHz, 40 dB SPL
MASKER 1 kHz, 60 dB SPL
MASKER 1 kHz, 80 dB SPL
SIMULTANEOUS
NONSIMULTANEOUS
60 dB SIMULTANEOUS
THRESHOLD (dB SPL)
PLACE OFFSET (percent)
SIGNAL FREQUENCY (kHz)
SIGNAL FREQUENCY (kHz)
SIGNAL FREQUENCY (kHz)

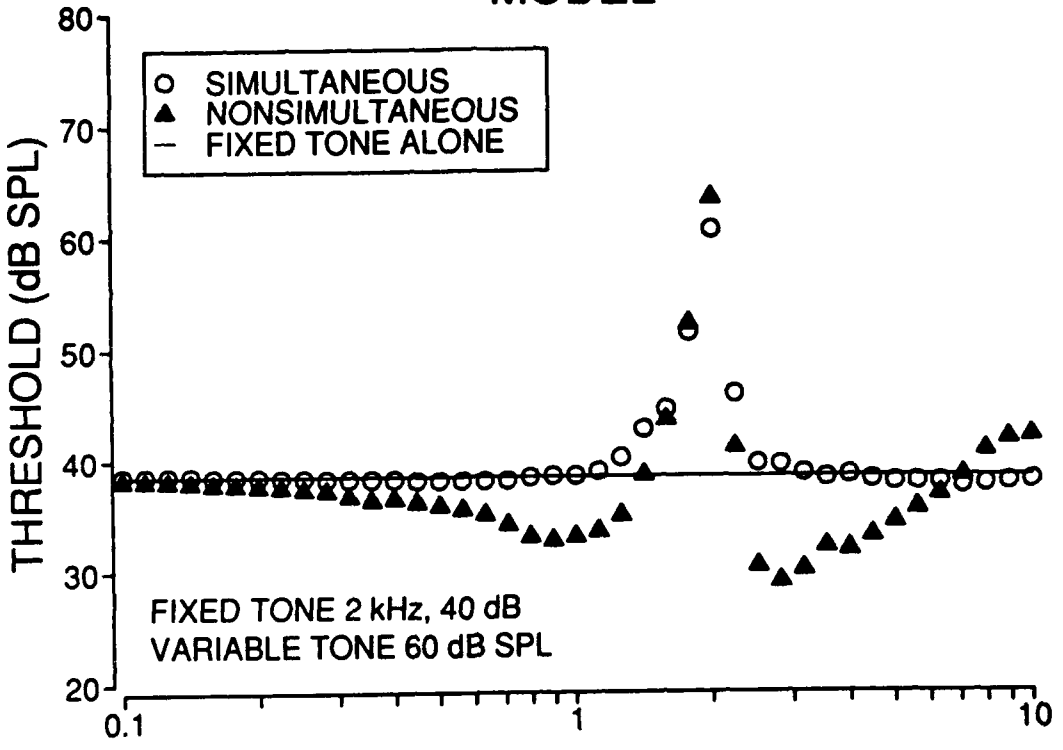
PSYCHOPHYSICAL TUNING CURVES



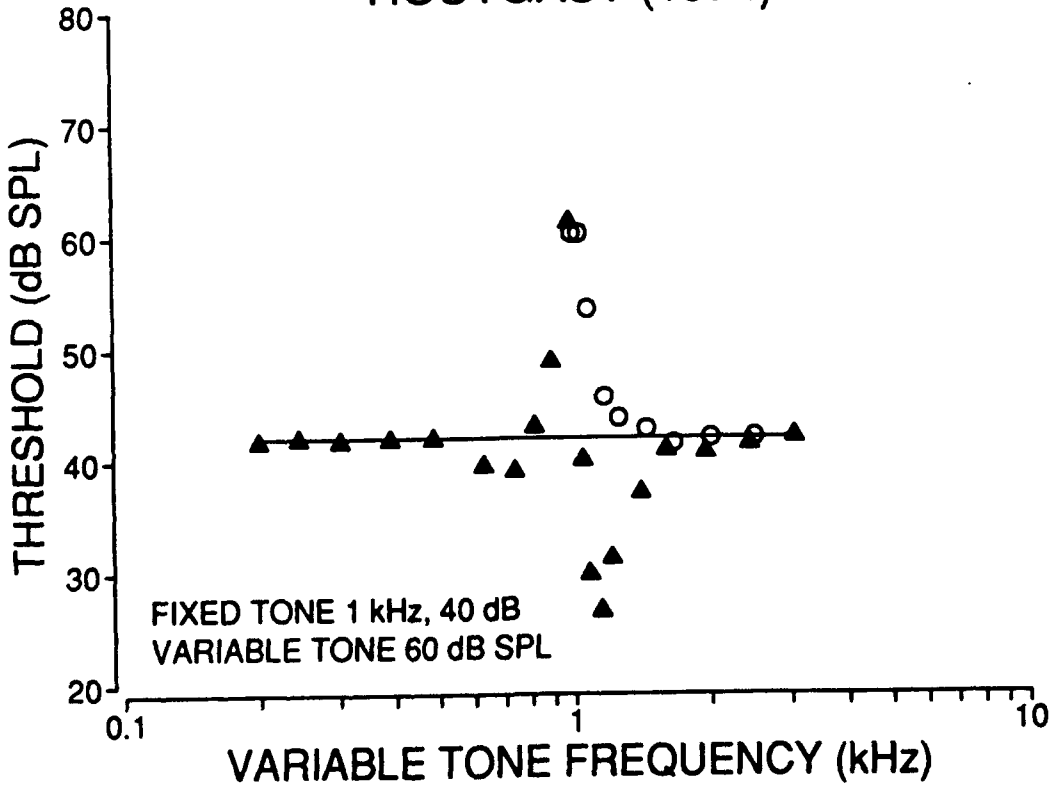

MODEL TUNING CURVES
100
80
60
40
20
0
-20
THRESHOLD (dB SPL)
NONSIMULTANEOUS PSYCHOPHYSICAL
SINGLE FIBER
0.1
1
10
MASKER FREQUENCY (kHz)

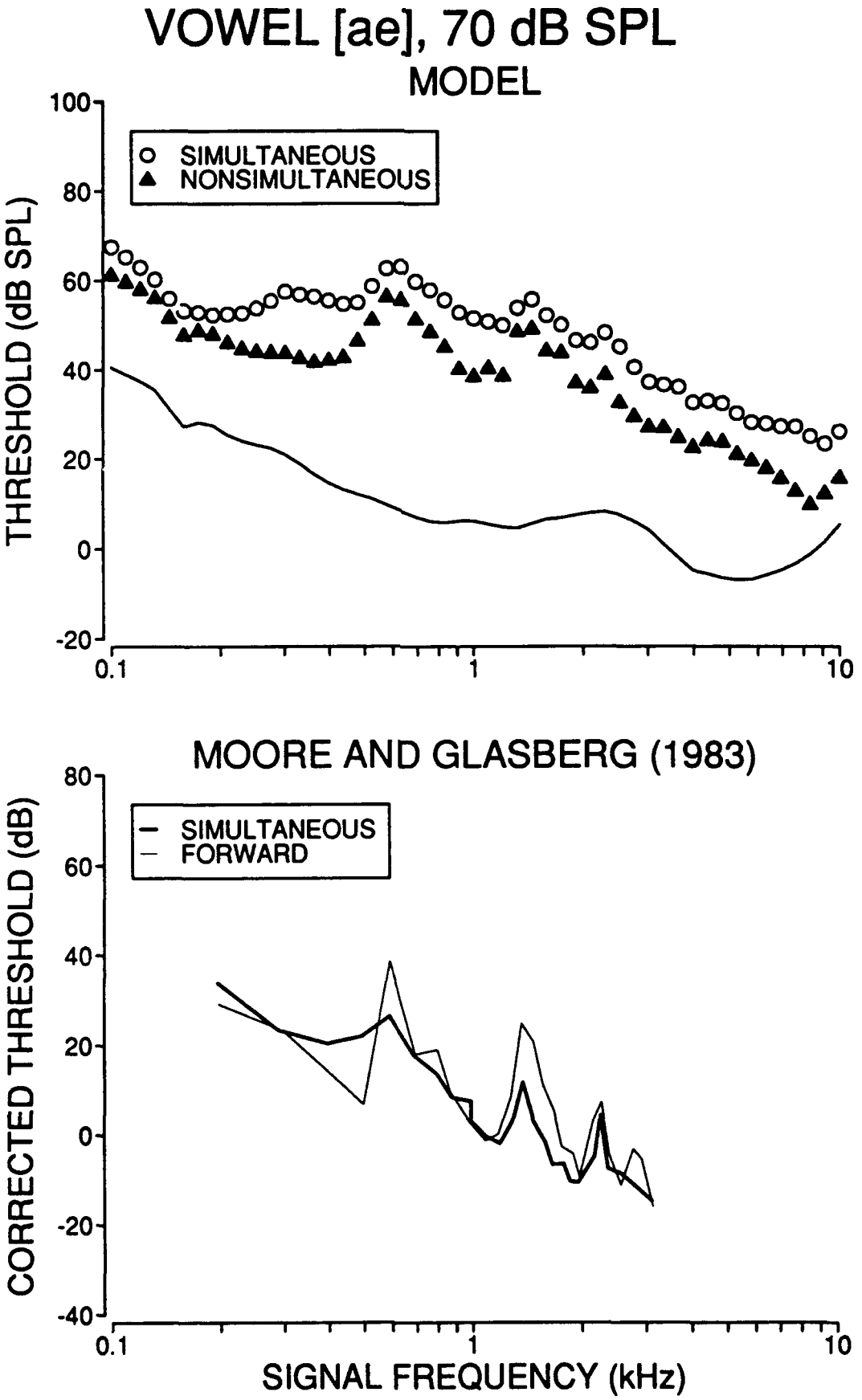
TWO-TONE UNMASKING

MODEL



HOUTGAST (1974)




VOWEL [ae], 70 dB SPL
MODEL
SIMULTANEOUS
NONSIMULTANEOUS
THRESHOLD (dB SPL)
100
80
60
40
20
0
-20
0.1
1
10
MOORE AND GLASBERG (1983)
SIMULTANEOUS
FORWARD
CORRECTED THRESHOLD (dB)
80
60
40
20
0
-20
-40
0.1
1
10
SIGNAL FREQUENCY (kHz)

ROLES OF SUPPRESSION IN MASKING

	NONSIMULTANEOUS MASKING	SIMULTANEOUS MASKING
WITHIN-MASKER SUPPRESSION	Unmasking	No unmasking
MASKER-TO-SIGNAL SUPPRESSION	None	Broadens masking pattern Upward spread of masking
SIGNAL-TO-MASKER SUPPRESSION	None	Improves signal detection in broadband maskers

INTERACTIONS BETWEEN NEUROPHYSIOLOGY AND SPEECH DISCRIMINATION

Donal G. Sinex, Lynn P. McDonald, and John B. Mott

University of California

Los Alamos National Laboratory

Physiology Group

Mail Stop M882

Los Alamos, NM 87545

505-665-1102

Presented at

ASA Spring Meeting

Syracuse, NY

May 1989

In this talk I will provide a brief discussion of our strategy for investigating the neural encoding of speech sounds. Then I will present results of experiments in which this strategy was applied to study the processing of voice onset time (VOT), an acoustic cue whose interesting perceptual properties extend to the chinchilla.

A. Approach

The availability of appropriate psychophysical data has been an essential part of our approach to studying neural mechanisms for the encoding of speech (i.e., Sinex and McDonald, 1988a, 1989) and other complex sounds (Sinex and Havey, 1984, 1986). Psychophysical experiments indicate what an animal can detect or resolve; they establish what neural computation the auditory system can carry out. By themselves, psychophysical results do not provide direct information about neurophysiological mechanisms; for example, knowing the magnitude of a frequency difference limen will not tell you whether it resulted from place coding or from a temporal analysis of frequency. If one is interested in the details of the neural computation, other kinds of measurements must be carried out. The utility of psychophysical data for us is that they help the investigator know what to expect from neural data. The significance of neurophysiological response patterns is not always straightforward. Even in a homogeneous structure like the auditory nerve, the pattern of responses is usually so complex that more than one interpretation of the results is possible; how does one choose among them? Prior knowledge of the animal's auditory capabilities can impose constraints on the interpretation of physiological observations. We assume that characteristics or details of the neural response that are correlated with psychophysical performance are most likely to be used by the CNS. Response properties that are independent of psychophysical observations are assumed to be less likely to contribute to that particular auditory function.

In some cases, more than one response property will correlate with psychophysical results. For example, Sinex and Havey (1986) found that the chinchilla's psychophysical masked thresholds (measured for the same stimulus conditions by Long and Miller [1981]) could be predicted from either average discharge rate or discharge synchrony. However, Sinex and Havey also noted that an FFT-based analysis of chinchilla auditory nerve fiber responses underestimated the behavioral thresholds.

Although Sinex and Havey could not state conclusively that detection at the masked threshold was mediated by one of these response properties, they were able to suggest that there is a constraint on the use of discharge synchrony. They concluded that if the CNS does use synchronized responses for detection, it must use an algorithm that is less sensitive than the processing that they had done. That conclusion, which argues for caution in interpreting the significance of synchronized responses to other sounds, could not have been drawn if the behavioral thresholds had not been available.

It should be noted that there are cases in which no aspect of the response may predict the psychophysical result. How should this result be interpreted? It can mean that the investigator has not yet found the relevant features of the neural code. It could also mean that the neural locus being studied contributes partially or not at all to the behavior (for example, some CNS structures do not measure interaural time even though it is known that animals can use it for sound localization). In these examples, the psychophysical data tell you that somehow the nervous system does a particular kind of computation and justifies additional experiments to determine where or how it is done.

I referred to "appropriate" psychophysical data in the introduction. Ideally, the psychophysical data will have been obtained from the same species as the neural data. The psychophysical task will have been as simple as possible - detection or discrimination - and data would be collected in procedures that minimize the influence of nonsensory factors such as memory. Also, differences in the stimuli used should be as small as possible. In practice, of course, these ideal criteria are not always met, especially when the experiments are conducted in different laboratories.

We do not expect that a correlation or a successful prediction based on, say, the responses of the auditory nerve minimizes the importance of CNS processing. Instead we think it helps identify those aspects of the peripheral response that form the basis for additional processing by the CNS, but the CNS remains of primary importance for perception. Also, we acknowledge that VOT, the topic of the rest of this talk, may be a special case, a speech feature whose processing by the general auditory system is particularly amenable to study. The role of the auditory periphery and the neural mechanisms that extract other speech features are probably quite different from what we have hypothesized to be true for VOT.

B. Experimental observations

I will now present some results from a study of the representation of VOT in responses of the chinchilla the auditory nerve. VOT is the delay between the release of an initial stop consonant and the onset of vocal fold activity or voicing. Voiced consonants such as /d/ or /g/ are produced with short VOTs, while voiceless consonants such as /t/ or /k/ are produced with long VOTs (Lisker and Abramson, 1964). Kuhl and Miller (1978) obtained psychophysical identification functions for synthesized VOT syllables from chinchillas and found them to be quite similar to identification functions obtained for English-speaking humans. Later, Kuhl (1981) found the temporal acuity for changes in VOT in the chinchilla subjects was greatest at the identification boundary; that is, the chinchillas exhibited a "phoneme boundary effect" (Wood, 1976). Several investigators had obtained comparable results from human listeners (Abramson and Lisker, 1970; Carney, et al., 1977; Soli, 1983; MacMillan, et al., 1988; however, see Kewley-Port, et al. 1988, and Watson and Kewley-Port, 1988, *for a different viewpoint*). These results obviously suggest that general auditory mechanisms underlie the effect, but as noted, psychophysical results cannot establish the nature or locus of the mechanism that is responsible. Therefore we measured the responses of single auditory nerve fibers in chinchillas to VOT syllables synthesized according to Kuhl and Miller's descriptions, to see what (if any) correlations might exist between peripheral response patterns and the nonmonotonic pattern of temporal acuity for VOT often observed psychophysically.

The trajectories of the first four formants (F1-F4) of the alveolar VOT continuum used in our experiments are shown in the first slide.

SLIDE 1

Following the release burst, the upper formants were aspirated and F1 was off until the onset of voicing (80 msec in this example). After VOT, all formants were voiced, with a fundamental frequency of 114 Hz. The syllables were always presented at 60 dB SPL.

The onset of voicing produced a rapid increase in low-frequency energy; spectra taken at the onset of the syllable and after the onset of voicing are shown in the next slide.

SLIDE 2

Because of this pattern, the responses of neurons with characteristic frequencies (CF)

below about 1.0 kHz exhibited the most distinctive responses to VOT. This talk will be limited to the average rate responses of those low-CF neurons.

C. Responses of single neurons to VOT

The responses of a typical low-CF neuron to several VOT syllables are shown in the next slide.

SLIDE 3

Each response profile exhibited a discharge rate increase whose latency generally increased with increasing VOT. In profiles such as these it appears that the response elicited by each VOT is quite distinct from the others, and there is no suggestion that the responses fall into two categories as the psychophysical results suggested. If the central nervous system (CNS) could isolate the responses of individual neurons such as this one, 10-msec VOT steps should be easily resolvable. Kuhl's psychophysical results, however, suggest that 10 msec steps are not usually resolvable by chinchillas. Since we have that information to guide us, we might conclude that information present in the peripheral auditory system is lost at some higher stage of processing. While that undoubtedly happens under some circumstances (when psychophysical procedures place high demands on memory, for example), I will argue that for the simple discrimination experiment conducted by Kuhl, that was probably not the case. Alternatively, it is possible that the analysis represented here is inconsistent in some significant way with the analysis done by the CNS. It is most likely that the CNS cannot monitor the response patterns of a single neuron as we have done here. To get a more realistic view of representation of the stimulus that is actually available to the CNS, we extended the analysis to examine responses of several neurons simultaneously, to ask what the spatial pattern of response to each VOT syllable is.

D. Responses of small populations of neurons to VOT

We compared the response profiles of groups of neurons with similar CFs to determine the precision with which they represent each VOT. The responses of nine low-frequency neurons from one animal to two VOTs separated by 10 msec are shown in the next slide.

SLIDE 4

The solid lines are responses to VOT= 30 msec, and the dashed lines are the responses of the same neurons to VOT= 40 msec. Each set is characterized by a discharge rate increase that closely matched the nominal stimulus VOT, as was true in the responses of the single neuron shown in the previous figure. While there was some variation across neurons in the response to each sound, these two sets of responses are easily separable by eye, because the rate increases occurred at different times. These two VOTs were easily resolvable by chinchillas in Kuhl's (1981) psychophysical experiment.

The same neurons' responses to two different VOTs in shown in the next slide.

SLIDE 5

The solid lines are responses to VOT= 60 msec, and the dashed lines are responses to VOT= 70 msec; the stimulus difference was 10 msec as before. The individual responses look different, however. The rate increase was gradual rather than abrupt, and it began prior to the nominal VOT. In this case, the responses overlap extensively, so that members of one group cannot easily be separated from the other. Kuhl's subjects could not resolve 10 msec changes in VOT when these sounds were presented.

Part of the response variability evident in this slide results from the fact that we have included neurons with different CFs. At any given CF, the threshold characteristics of neurons also contribute to the result. However, it is still the case that the same CF and threshold variation does not introduce the same amount of variability into the responses to 30 and 40 msec VOT stimuli. Spectral cues for the middle VOTs (30 and 40 msec) occur with abrupt rise times so that they recruit neurons with different thresholds synchronously, marking VOT with great precision. In contrast, sounds that rise gradually, as is true for the long VOTs, recruit sensitive neurons early and less-sensitive neurons later, reducing the temporal precision of the population response. This is illustrated in the responses of two neurons with the same CF but different thresholds and spontaneous rates, shown in next slide.

SLIDE 6

In response to VOTs of 30 and 40 msec, the response profiles were temporally similar although of different magnitude; the latency difference was 8 msec. For long VOTs,

SLIDE 7

the temporal patterns as well as the response magnitudes were different. For VOTs of 60 and 70 msec, the latency difference was 32 msec, a factor of 4 larger than for the 30-40 msec case.

We have summarized response patterns such as these by defining a response latency from each individual neuron's discharge rate profile. We then calculate the mean and standard deviation of latency across neurons; examples are shown in the next slide.

SLIDE 8

Each curve represents the responses of all the low-CF neurons recorded from an individual animal. The pattern across animals was consistent. Mean latency [called m(L) in the figure] increased approximately in proportion to the stimulus VOT, but the standard deviation s(L) was smallest for VOTs of 30-40 msec.

E. Quantitative prediction of temporal acuity for VOT

The data in the previous slide indicate that the information passed to the CNS by populations of neurons is imperfect, with limits that are apparently consistent with the pattern of psychophysical temporal acuity. To make a more quantitative test of this possibility, we generated a model for the pairwise discriminability of VOT syllables (Sinex and McDonald, 1988b,c; Sinex, et al., 1989). The model is based on Signal Detection Theory (Green and Swets, 1974) and assumes that the discrimination of VOT by the CNS is limited only by the response variability of primary neurons. This is similar to the notion of sensory variance discussed by MacMillan, et al. (1988), except that those authors assume the sensory variance is independent of the stimulus. Our measurements of course indicate a nonmonotonic dependence on VOT. We have simulated thresholds for the same conditions tested by Kuhl (1981) and show the result in the next slide.

SLIDE 9

The pattern of neural thresholds (solid lines) was similar to the psychophysical thresholds (dashed lines): V-shaped, with a minimum occurring at 30 or 40 msec, depending upon the direction of stimulus change. The quantitative agreement was also good, considering that the stimuli were not identical.

F. What is the mechanism?

What properties of the auditory periphery account for these findings? The result cannot reflect the operation of feature detecting circuits, since no circuits involving lateral connections between neurons are known to exist in the auditory periphery. Rather, the observed patterns of variability arise through fairly simple interactions between the low-frequency spectra of the VOT syllables and the tuning and threshold characteristics of the neurons. The temporal pattern of low-frequency spectral change is compared to the rate responses of low-frequency neurons in the next slide.

SLIDE 10

The neural responses at the top of the figure were shown earlier as a separate slide. Note the rapid rise of discharge rate and the tight clustering of responses to each stimulus. The lower part of this slide looks similar but was generated by analysis of the stimulus waveforms with simple linear filters, shaped approximately like chinchilla tuning curves; the slide shows the temporal pattern of spectral change for the same two VOTs. The lines represent the energy in a syllable passed by tuning curve filters with center frequencies between 300-900 Hz; 0 dB on the ordinate represents the output generated by a signal at the threshold of the tuning curve that served as the model for the filters, so that outputs to sub-threshold inputs are not shown. The solid lines are measurements for VOT= 30 msec and the dashed lines are for VOT= 40 msec. This and other related analyses indicate that the latency of the first response to VOT (and to a lesser extent the general temporal pattern of the neural response) can be approximated by linear filtering.

The next slide

SLIDE 11

presents spectral analysis and neural responses for VOTs of 60 and 70 msec. In this case, the rise time of the neural response was gradual, and the responses elicited by the two sounds were not easily separated. When the responses are compared to the spectral analyses obtained from the same linear filters as before, it can be seen that the neurons began to respond to spectral cues that were present before the onset of voicing. These cues did not distinguish between the two VOTs. After VOT, when the filter output is greater than 20 dB above threshold, a VOT-dependent acoustic cue was present. The neurons do exhibit a small response peak whose latency reflects VOT, but this peak is

small relative to the discharge rates that had already been elicited. The peak is probably much less salient to the CNS than the clear discharge rate peaks seen in the previous slide.

G. Conclusions

The response patterns elicited in the peripheral auditory system by VOT syllables appear to contribute to the nonmonotonic temporal acuity for VOT observed in most psychophysical experiments. Response variability over a fixed array of low-frequency neurons was smallest for VOTs of 30-40 msec, leading to a prediction of increased temporal acuity for those syllables. Alternatively, one could describe the results in terms of the spatial extent over which the onset of voicing elicited a synchronous (and therefore temporally precise) response. For middle VOTs, many neurons signal the onset of voicing with a robust response increase. For longer and shorter VOTs, the population of neurons whose discharge rates increase at any one time is smaller. Either way, the representation of the onset of voicing elicited by VOTs in the 30-40 msec range appears to be much more salient, more easily processed by the CNS.

In the future we hope to determine in more detail how groups of neurons respond to VOT syllables. For example, responses to other VOT continua and effects of stimulus level have not been adequately explored. The simulations using tuning-curve based filters mentioned above will be used for at least some of these investigations. Because the VOT response is largely determined by the properties of the neuron at threshold, the simple linear model will be useful for this stimulus.

In closing, it should be stated once again that these results leave a very important role in VOT processing for the CNS. The statistical model that we used to predict Kuhl's data is in effect a hypothesis about the performance that the CNS is capable of, given the nature of the signals that it receives. The question of how the CNS processes the variable responses of auditory nerve fibers is yet to be answered.

References

Abramson and Lisker (1970). Proc. 6th Intl. Congress of Phonetic Sci., Prague (Academic), 569-573.

Carney, et al., (1977). J. Acoust. Soc. Am., 62, 961-970.

Green and Swets (1974). Signal Detection Theory and Psychophysics (Krieger, Huntington, NY).

Kewley-Port, et al. (1988). J. Acoust. Soc. Am., 83, 1133-1145.

Kuhl (1981). J. Acoust. Soc. Am., 70, 340-349.

Kuhl and Miller (1978). J. Acoust. Soc. Am., 63, 905-917

Lisker and Abramson (1964). Word, 20, 384-422.

Long and Miller (1981). Hear. Res., 4, 279-285.

MacMillan, et al. (1988). J. Acoust. Soc. Am., 84, 1262-1280.

Sinex and Havey (1984). Hear. Res., 13, 285-292.

Sinex and Havey (1986). J. Neurophysiol., 56, 1763-1780.

Sinex and McDonald (1988a). J. Acoust. Soc. Am., 83, 1817-1823.

Sinex and McDonald (1988b). Meeting of ARO, February 1988.

Sinex and McDonald (1988c). Meeting of ASA, May 1988.

Sinex and McDonald (1989). J. Acoust. Soc. Am., 85, 1995-2004.

Sinex, et al., (1989). Meeting of ARO, February 1989

Soli (1983). J. Acoust. Soc. Am., 73, 2150-2165.

Watson and Kewley-Port (1988). J. Acoust. Soc. Am., 84, 2266-2270.

Wood (1976). J. Acoust. Soc. Am., 60, 1381-

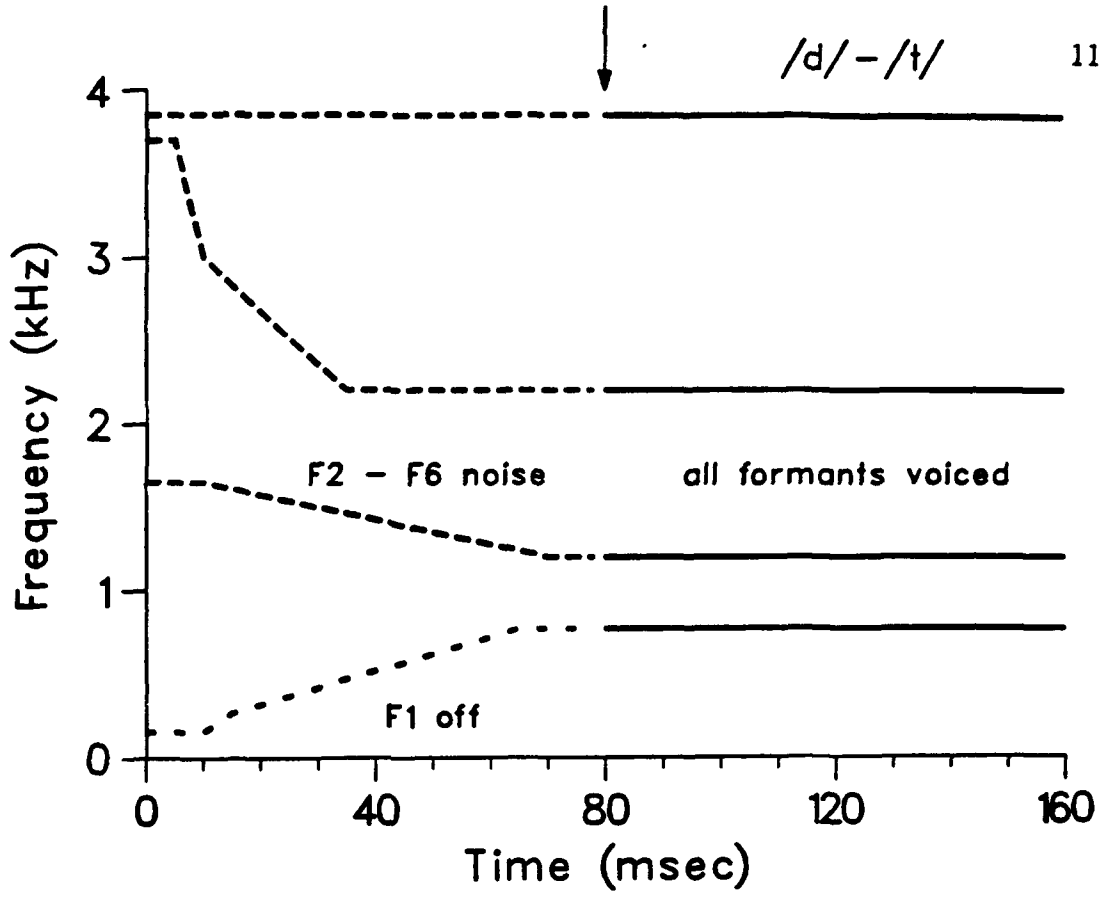
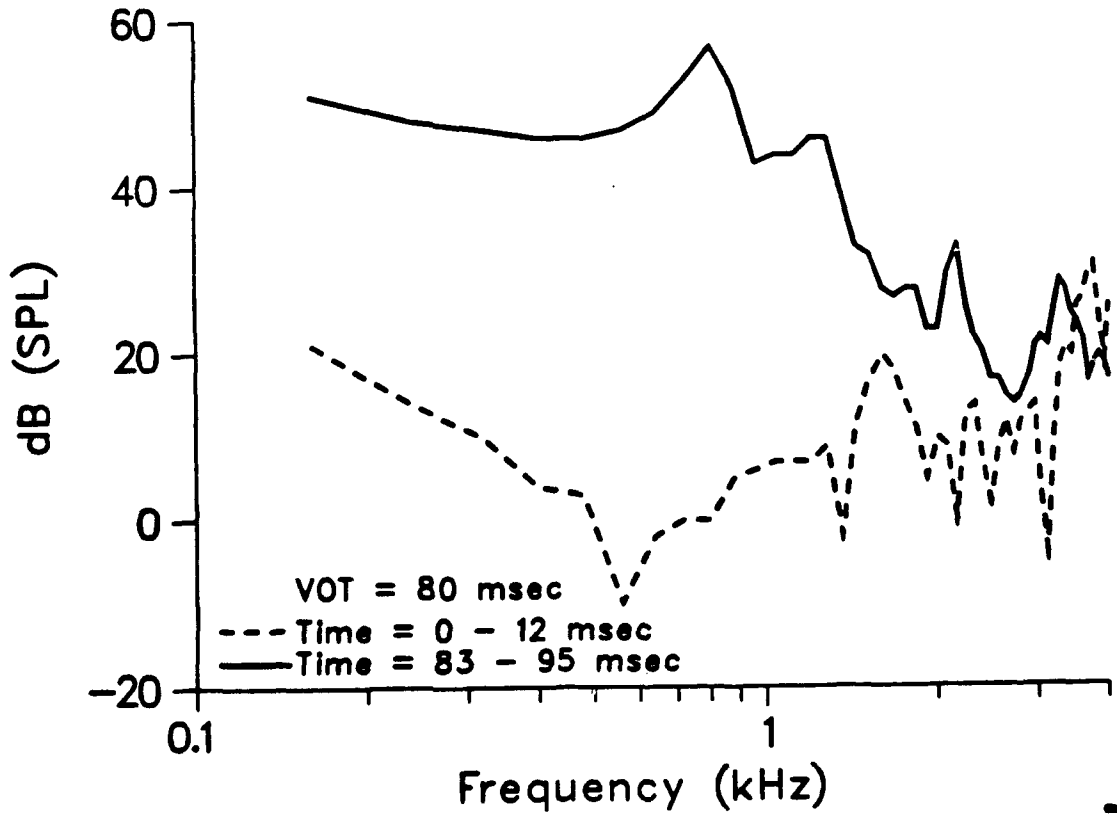
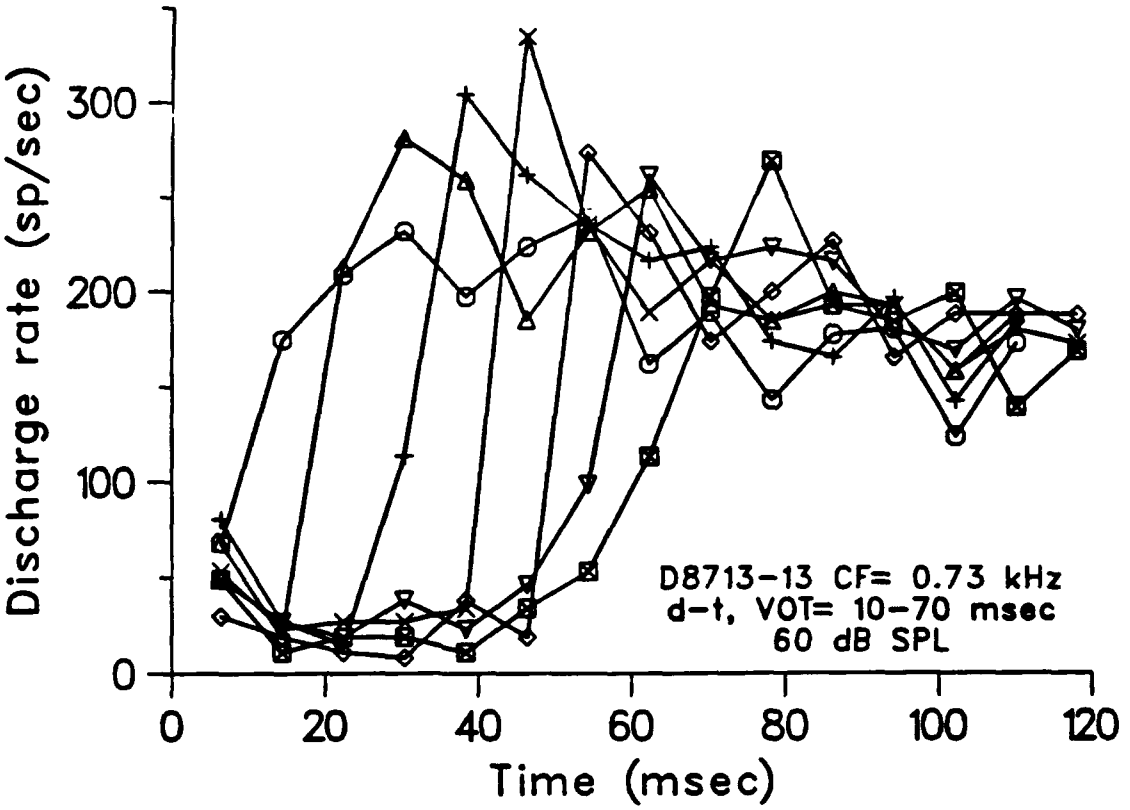
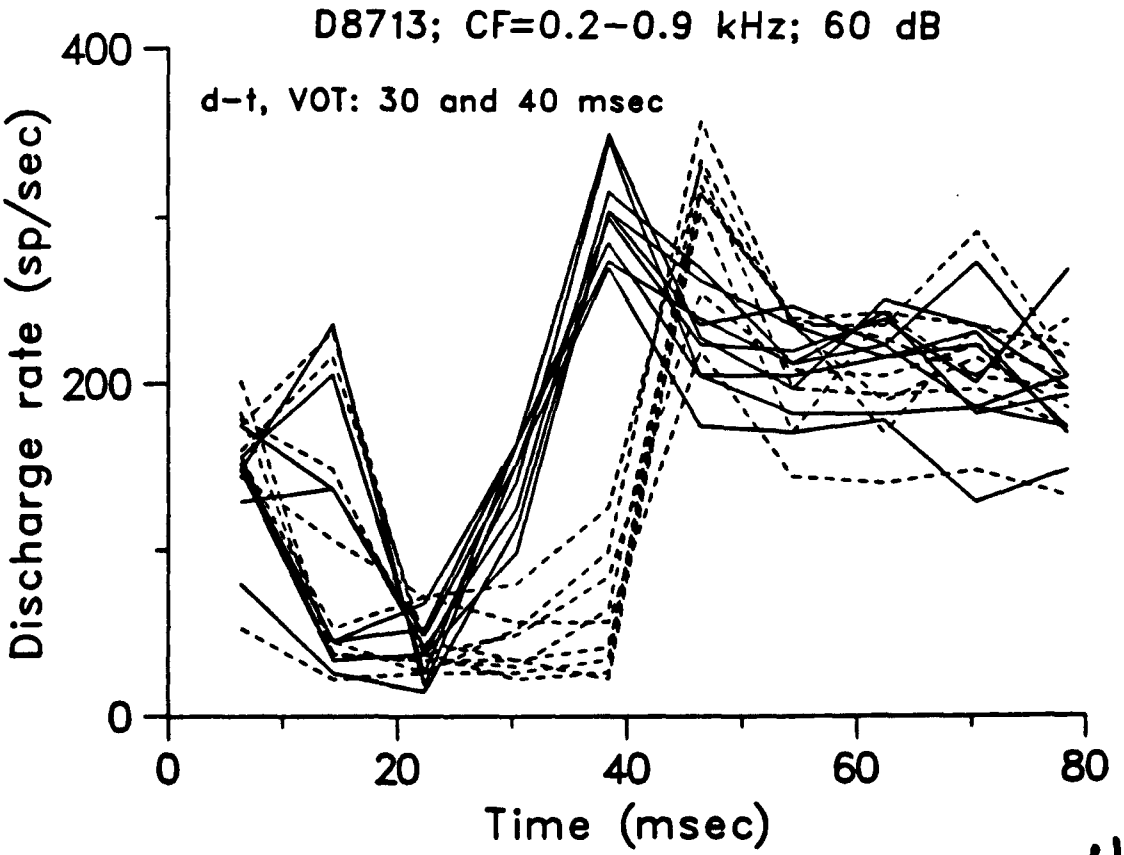

/d/ - /t/
Frequency (kHz)
4
3
2
1
0
F2 - F6 noise
all formants voiced
F1 off
0
40
80
120
160
Time (msec)

FIG. 1

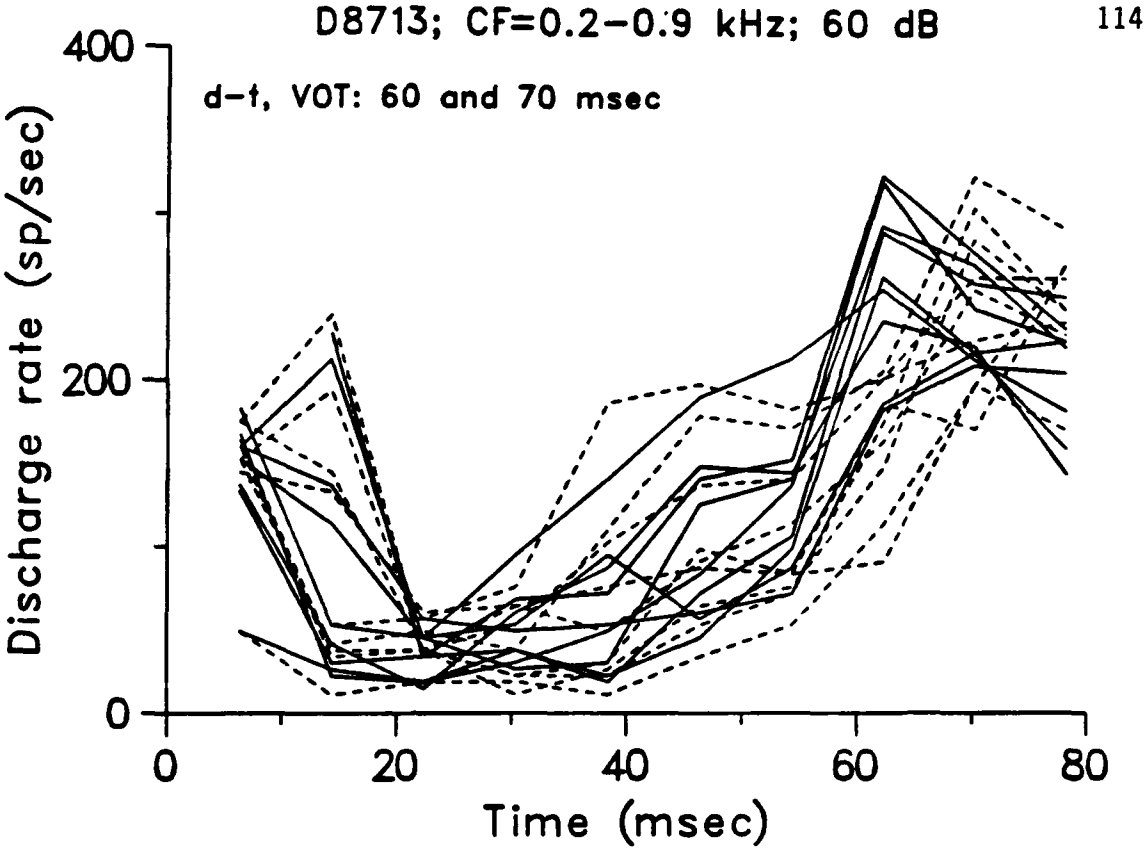

dB (SPL)
60
40
20
0
-20
VOT = 80 msec
---Time = 0 - 12 msec
——Time = 83 - 95 msec
0.1
1
Frequency (kHz)


Discharge rate (sp/sec)
300
200
100
0
0
20
40
60
80
100
120
Time (msec)
D8713-13 CF= 0.73 kHz
d-t, VOT= 10-70 msec
60 dB SPL

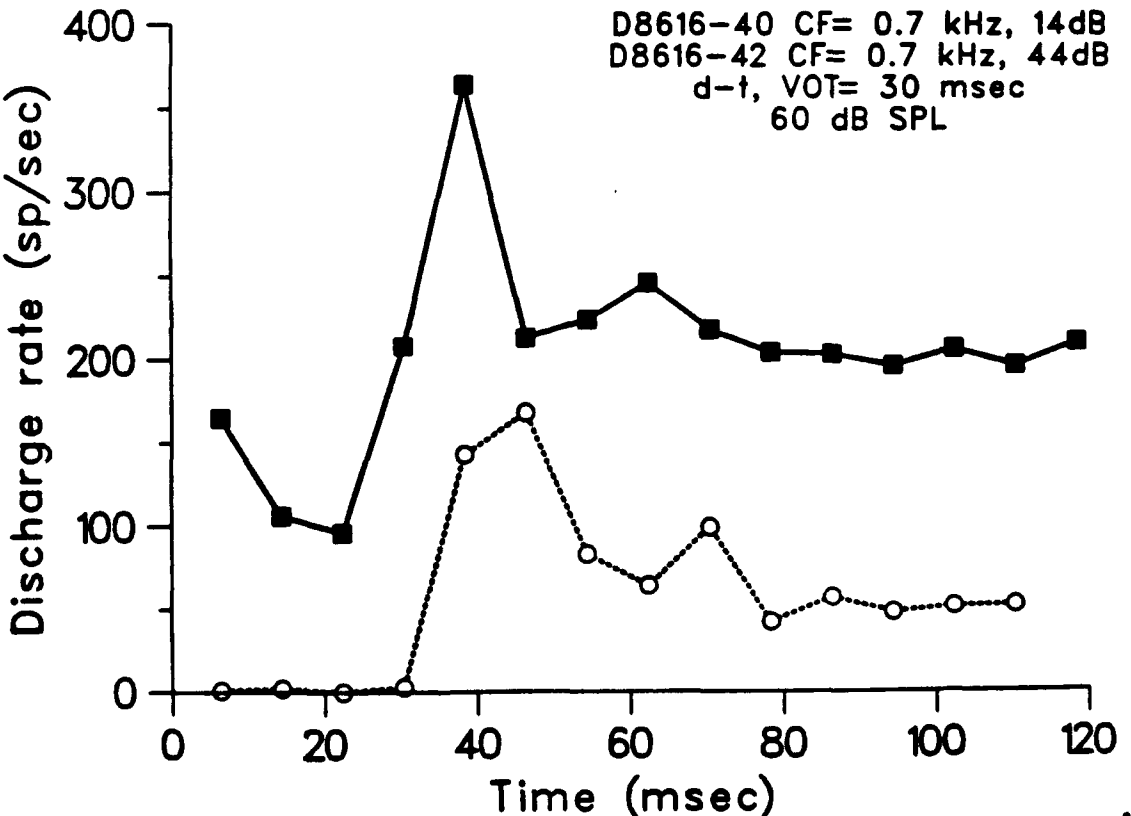
3


D8713; CF=0.2-0.9 kHz; 60 dB
d-t, VOT: 30 and 40 msec
Discharge rate (sp/sec)
400
200
0
0
20
40
60
80
Time (msec)

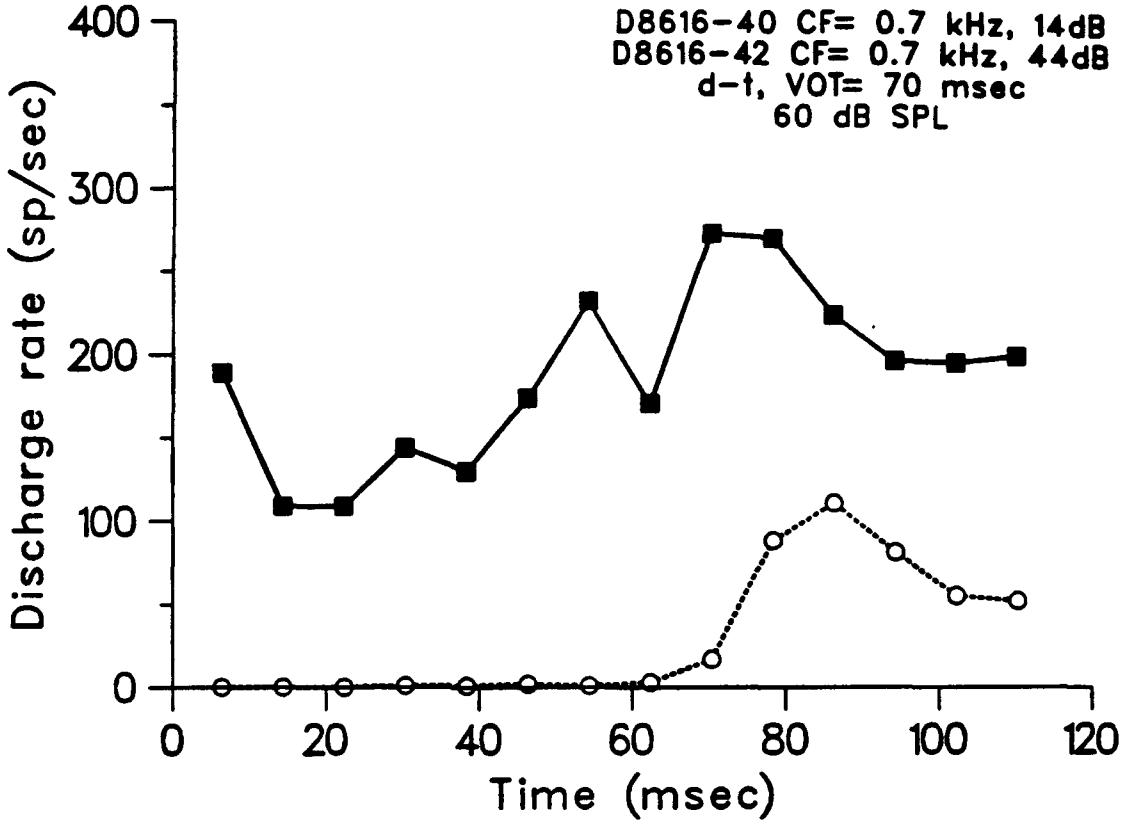
4


D8713; CF=0.2–0.9 kHz; 60 dB
d–t, VOT: 60 and 70 msec
Discharge rate (sp/sec)
400
200
0
0
20
40
60
80
Time (msec)

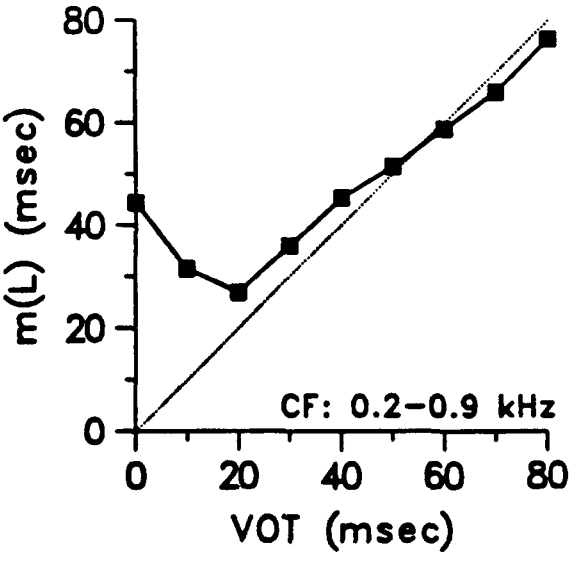
5

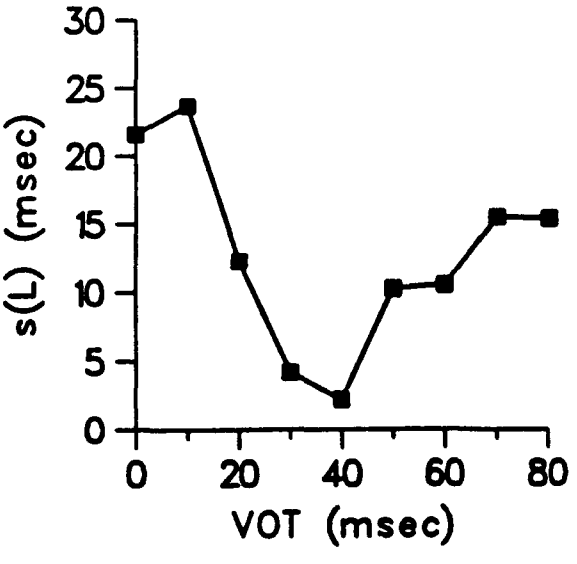

D8616–40 CF= 0.7 kHz, 14dB
D8616–42 CF= 0.7 kHz, 44dB
d–t, VOT= 30 msec
60 dB SPL
Discharge rate (sp/sec)
400
300
200
100
0
0
20
40
60
80
100
120
Time (msec)

6

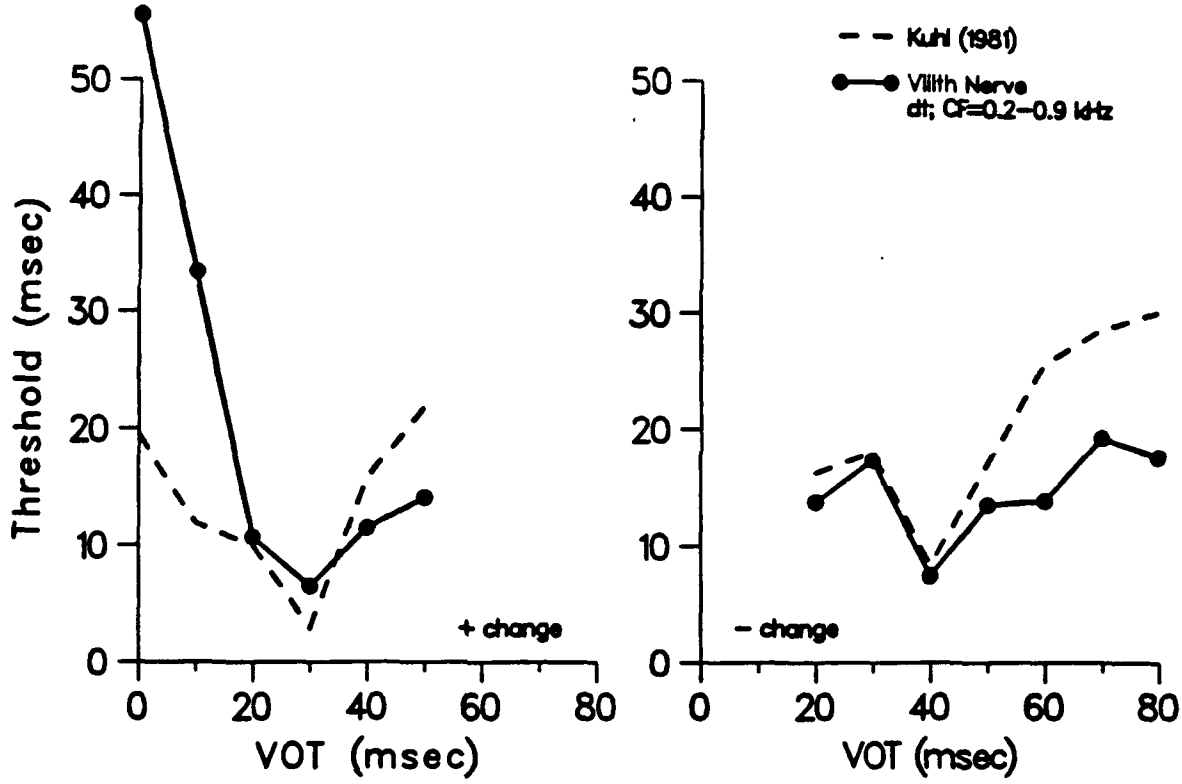

D8616-40 CF= 0.7 kHz, 14dB
D8616-42 CF= 0.7 kHz, 44dB
d-t, VOT= 70 msec
60 dB SPL
Discharge rate (sp/sec)
400
300
200
100
0
0
20
40
60
80
100
120
Time (msec)

7

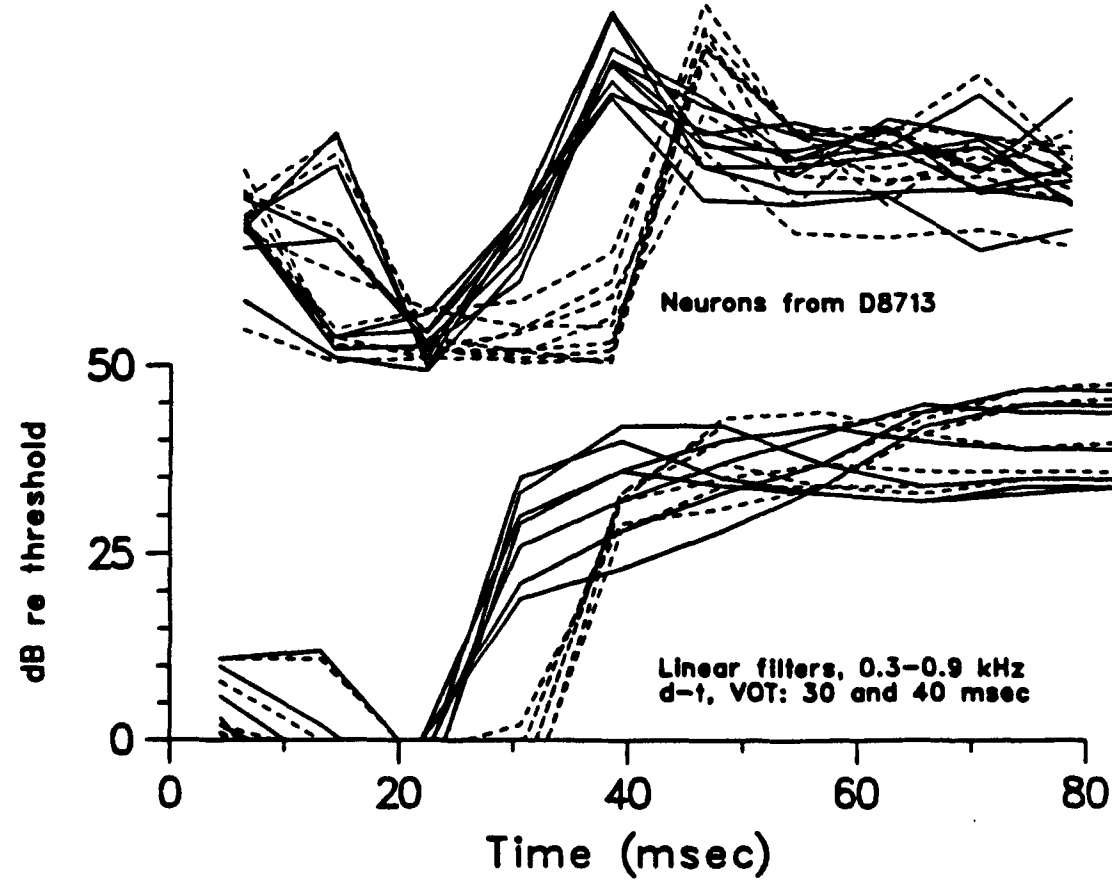

m(L) (msec)
80
60
40
20
0
CF: 0.2-0.9 kHz
0
20
40
60
80
VOT (msec)

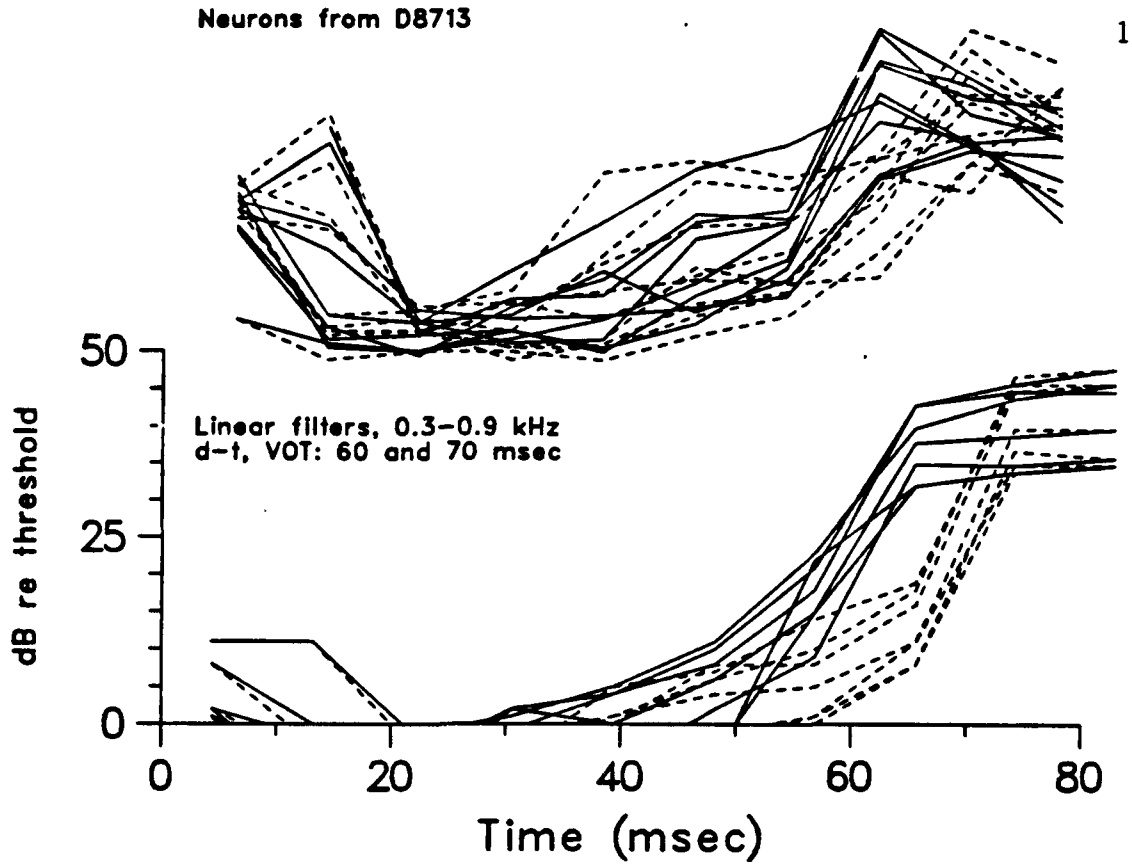

s(L) (msec)
30
25
20
15
10
5
0
0
20
40
60
80
VOT (msec)

8


Kuhl (1981)
VIIIth Nerve
d1; CF=0.2–0.9 kHz
Threshold (msec)
50
40
30
20
10
0
0
20
40
60
80
VOT (msec)
+ change
– change

9


Neurons from D8713
dB re threshold
50
25
0
0
20
40
60
80
Time (msec)
Linear filters, 0.3–0.9 kHz
d–t, VOT: 30 and 40 msec


Neurons from D8713
Linear filters, 0.3–0.9 kHz
d–t, VOT: 60 and 70 msec
dB re threshold
50
25
0
0
20
40
60
80
Time (msec)

11

Problems and Opportunities in Extending Psychophysical/Physiological Correlation into the Central Nervous System

Eric D. Young

Dept. of Biomedical Engineering and Center for Hearing Sciences,
Johns Hopkins School of Medicine

Presented at the 117th meeting of the Acoustical Society of America in Syracuse, NY, May 23, 1989

Introduction

Over the past 30 years, significant progress has been made in accounting for the properties of auditory psychophysics by using the response characteristics of auditory nerve (AN) fibers. Analysis of this problem has almost always been based on the approach of William Siebert (53), in which detection and estimation theory are applied to AN spike trains or models of AN spike trains in order to predict some psychophysical measure such as a jnd or a masked threshold.

Over the same time period, substantial progress has been made in characterizing the morphology and response properties of neurons in the central auditory system, especially in the first central processing center, the cochlear nucleus (CN; for reviews of this work, see refs 14, 26, 62, 65). The availability of information on the morphological organization of the CN and on the basic response characteristics of its neurons allows the program of psychophysical/physiological correlation to be extended into the CNS. This is an exciting opportunity, because it allows us to begin to study the information processing activities of the neural circuits of the auditory system directly. However, there are a number of new difficulties in working at the level of the CN that are not encountered when working in the AN. These difficulties are discussed in this paper, along with some possible approaches to their circumvention.

Figure 1 is a cartoon summary of the morphology of the cat CN. This figure illustrates the first problem that arises in considering psychophysical/physiological correlation in the CNS: at the CN the auditory pathway breaks up into at least five parallel channels. Whereas AN fibers form a relatively homogeneous array of elements differing from one another mainly in quantitative ways, such as tuning and bandwidth (18,31) or spontaneous rate (34), the principal cells of the CN have diverse characteristics. These cells differ from one another in almost every way that neurons can differ. Each cell type has its own characteristic structure (8,35), pattern of inputs from the auditory nerve (9, 13, 28, 35, 39, 58, 59), membrane and integrative properties (25,37), and pattern of

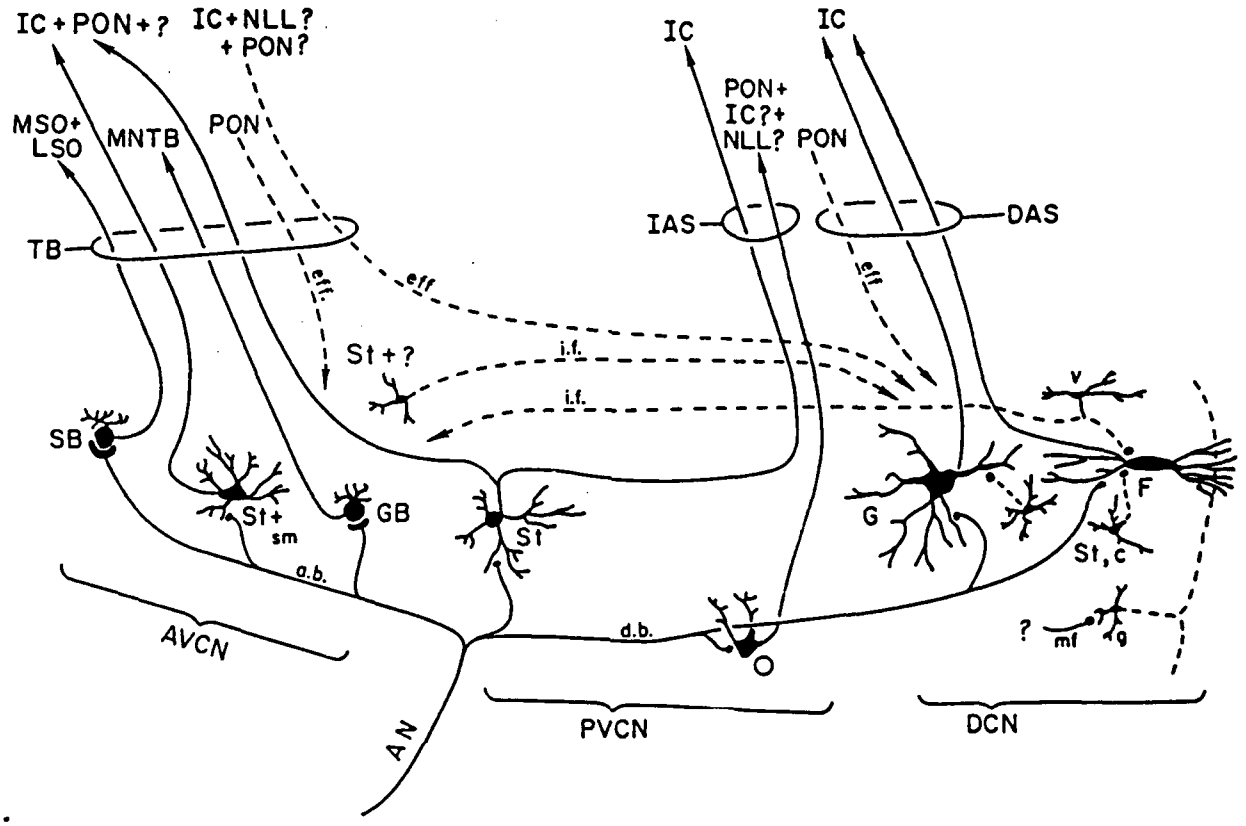


Figure 1 Schematic summary of the major subsystems of the cochlear nucleus (CN). Inputs to the CN come from auditory nerve (AN; 35) and from efferent fibers (eff) of higher centers (14). synaptic terminals are also formed by local interneurons, mainly in the dorsal CN (DCN; g, v, st, c) and by intranuclear association fibers (i.f.) that interconnect the subdivisions of the CN (2,35). Five principal cell classes (8,38) are shown. Bushy cells (SB, GB) receive large synaptic terminals from AN fibers on their somas and have limited dendritic trees (13,35,48,58,59). Their axons travel through the trapezoid body (TB) to the principal nuclei of the superior olive (MSO, LSO, MNTB; 11,60,61). Stellate cells (St, also called multipolar cells in some circumstances) have long dendrites which receive bouton terminals from AN and other sources; some St cells receive few or no terminals on their somata (9,59). St axons travel to the DCN and through TB and intermediate acoustic stria (IAS) to the periolivary complex (PON), inferior colliculus (IC), and contralateral CN (1,2,10,12,61). Octopus cells (O) receive massive AN input on soma and proximal dendrites (28); O axons project through IAS to several destinations (61). Principal cells of the DCN (G, F) receive inputs through intricate interneuronal circuits (g, c, st, v; 29,35,39) and are not discussed further.

projection to higher order nuclei in the brainstem (14,61). The complexity of morphological organization ranges from the apparently simple endbulb-of-Held synapse, which constitutes a massive input from a few AN fibers to a bushy cell (SB or GB in Fig. 1; 35,48,59), to the complex neuropil of the dorsal cochlear nucleus (DCN; 29,35,39).

The diversity of principal cell systems in the CN implies that it is necessary to analyze these systems separately in considering possible psychophysical/physiological correlations. This inference is supported by data on the response properties of CN neurons, summarized below. The output subsystems of the CN convey very different representations of the acoustic environment and we must assume, in approaching the auditory system, that each subsystem participates in audition in a different way. In this paper, the issues that arise in treating the different subsystems of the CN are discussed using examples from the ventral cochlear nucleus (VCN).

Response Properties of VCN Neurons

Fig. 2 shows examples of the three most common response types in the VCN. The plots in this figure are PST histograms of responses to 25 ms tone bursts at the best frequencies of the neurons. Each column shows a major response type, within which there are subtypes. In the VCN, unit types are usually named for their PST histogram response type. However, the classification scheme based on PST histograms is more general, in that a variety of other response characteristics follow once PST histogram type has been defined (3, 7, 30, 42, 64). Most important, it has been shown that there is a morphological correlate of PST response type, at least for primarylike units, which are recorded from bushy cells, and for chopper units, which are recorded from stellate cells (43,47,54). The morphological correlate of onset units is uncertain, although it is clear that some onset types are recorded from bushy cells (47) and others from octopus cells and large multipolar cells of the posterior VCN (20,43,45).

The inset of Fig. 2 shows spike trains from a primarylike and a chopper unit (46). These examples illustrate the striking difference between these two unit types in discharge regularity. The response of the primarylike unit is irregular, in that the intervals between spikes vary in an apparently random fashion and the pattern of response to successive stimuli is not repeatable. By contrast, the response of the chopper unit is regular, in that successive interspike intervals are similar and responses to successive stimuli are repeatable.

Fig. 3 shows the population distribution of regularity for primarylike and chopper units (64). This graph shows the regularity of discharge of a population of units in unanesthetized decerebrate cats, plotted as the standard deviation of interspike intervals (ordinate) versus the mean interspike

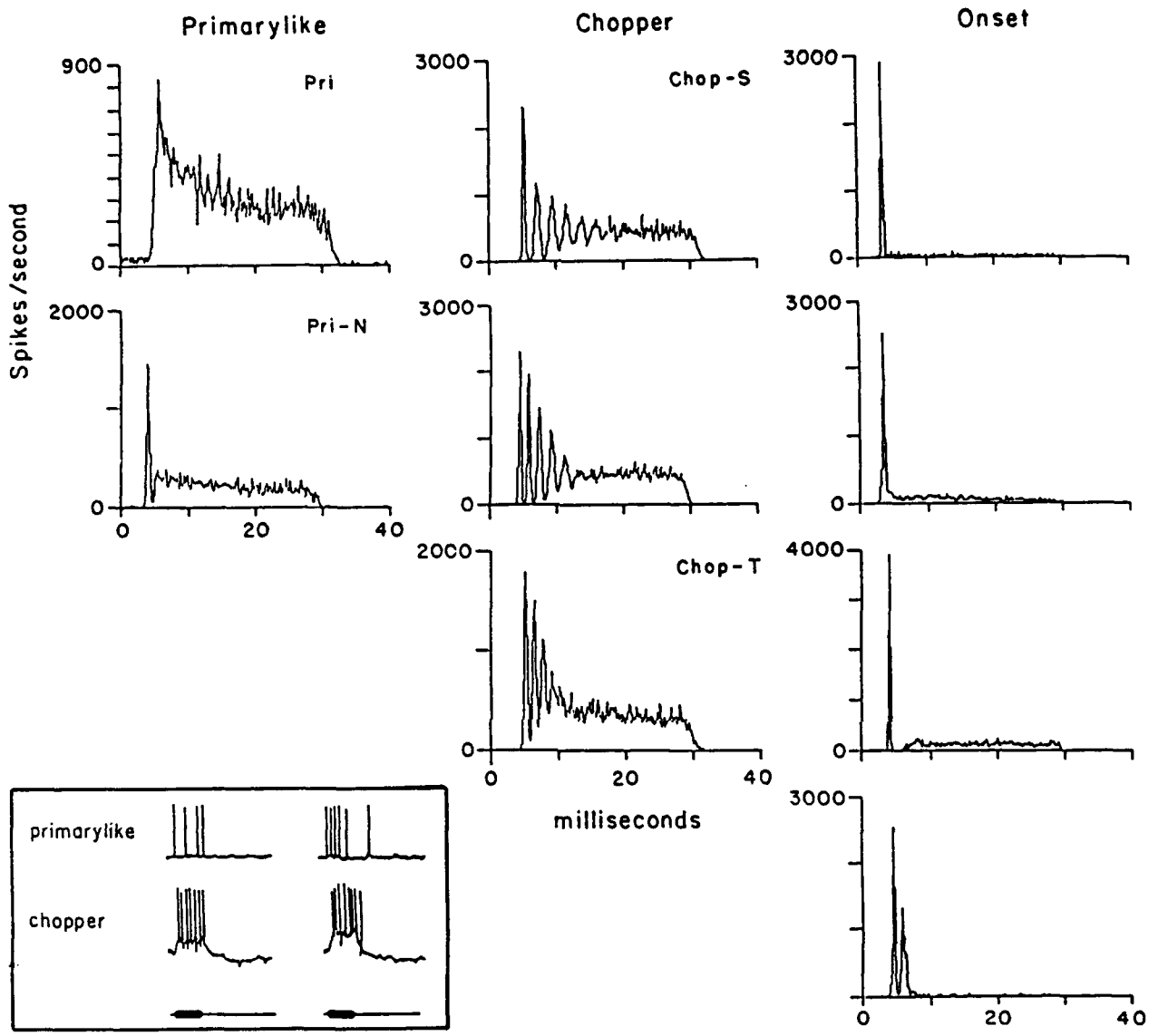


Figure 2 Examples of the PST histograms of the three major response types in VCN; PSTs are constructed from responses to 25 ms best-frequency tone bursts. Tones are on during the first 25 ms of the abscissa. Pri units give responses similar to those of AN fibers: Pri-N (for primarylike with notch) units are characterized by a sharp onset response followed by a short pause, or notch. Chopper units fire at regular preferred times following stimulus onset; times are not related to phase locking. Chop-S and Chop-T units are distinguished by regularity of discharge (see Fig. 3; 3,7,64). Onset units give one spike at stimulus onset, followed by nothing or a low discharge rate. These response types are based on analyses by Pfeiffer (41) and Bourk (7); the examples are taken from Blackburn and Sachs (3).

The inset shows spike trains from intracellular recordings of a primarylike and chopper unit (redrawn from ref 46).

interval (abscissa). The more regular a unit's discharge, the smaller the standard deviation will be. However, because standard deviation tends to grow with mean interval (22), it is the ratio of standard deviation to mean interval (called *coefficient of variation* or CV) that is usually used as the measure of regularity. Each point represents data from one stimulus condition in one unit, and there may be up to four data points from a particular unit.

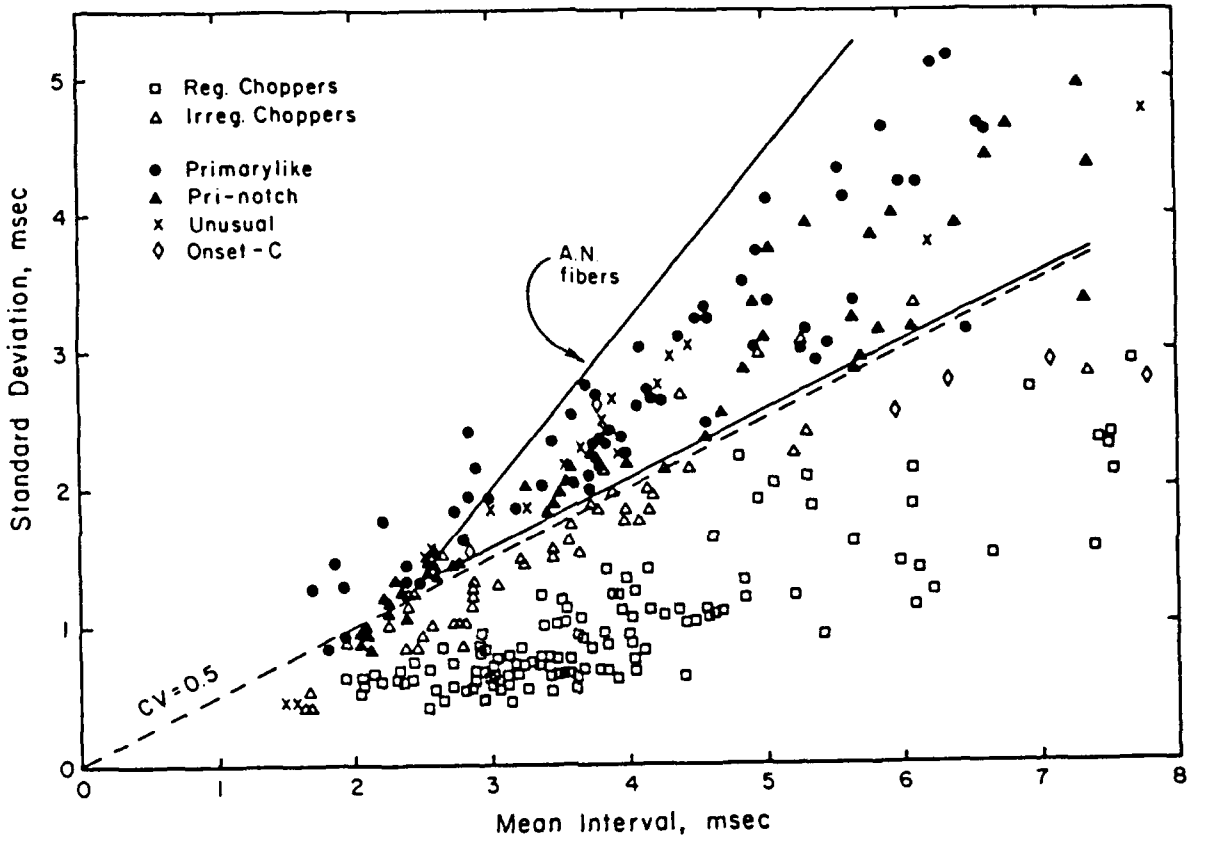


Figure 3 Standard deviation of interspike intervals plotted versus mean interspike interval for primarylike and chopper units (see legend). Unusual units (X) are difficult to classify units which behave in all ways like primarylike units. Reg. Choppers are Chop-S units and Irreg. Choppers are Chop-T units. Data are from responses to best-frequency tones from 10 to 40 dB re threshold; statistics computed from intervals beginning between 12 and 20 ms after the onset of 25 ms tone bursts. Reproduced from ref 64.

Chopper units are shown in Fig. 3 as open symbols and primarylike units as filled symbols. Note that most chopper units are regular, in that their CVs are less than 0.5, and primarylike units are irregular, in that their CVs are greater than 0.5. Plots of similar data for AN fibers yield points which fall within the V-shaped area marked "AN fibers". Fig. 3 shows that primarylike units have regularity comparable to that of AN fibers and that there is a sharp distinction between primarylike and chopper units in terms of regularity. A subsequent study with a larger number of units in anesthetized animals

(3) has given similar results, except that a larger population of irregular chopper units (chop-T units) was found and the overlap of chop-T units and primarylike units was larger than in Fig. 3. Nevertheless, it is clear that the most regularly discharging units in the CN are chopper units (especially chop-S units) and that primarylike units are irregular.

Detection of Acoustic Stimuli through AN and CN Rate Changes

The differences in regularity shown in Fig. 3 imply differences in the precision with which primarylike and chopper units carry information about the acoustic stimulus. The difference can be defined in terms of the theory of signal detection (23) in which the detectability d' of a change ΔR in some quantity, such as the response R of a neuron, is measured by the size of the change relative to the standard deviation σ_R of the quantity:

$$d' = \frac{\Delta R}{\sigma_R}$$

The probability of correct detection of ΔR depends both on the size of the change (detection probability increases with ΔR) and the noisiness or variability of R (detection probability decreases with σ_R). d' is a measure of detectability in that d' increases as the probability of detection of ΔR increases in a two-alternative forced choice experiment (23).

Because chopper units are more regular than primarylike units, σ_R should be smaller in chopper units for most response measures. Therefore, if a change in the stimulus induces the same response change in both chopper and primarylike units, the response change of the chopper should be more detectable than that of the primarylike unit. This difference can be illustrated by considering the case of detection of a tone in noise.

Fig. 4 shows data summarizing the responses of an AN fiber and a CN unit to best-frequency tones in the presence of background noise. The stimulus is shown schematically at the top of the figure and rate versus level functions for an AN fiber and VCN unit are shown below. The rate versus level functions show discharge rate during the tone bursts as a function of tone level, with background noise level as a parameter. Three effects of the noise on responses to the tones are seen (16,19). First, there is a response to the noise alone that increases the minimum discharge rate at low tone levels; second, there is a decrease in the maximum discharge rate (the saturation rate at high tone levels) which is caused by adaptation to the steady noise stimulus; and third, there is a rightward shift in the dynamic region of the unit's response which is caused by suppression of the response to the tone by the noise (16). These three effects are qualitatively similar in AN fibers and CN units,

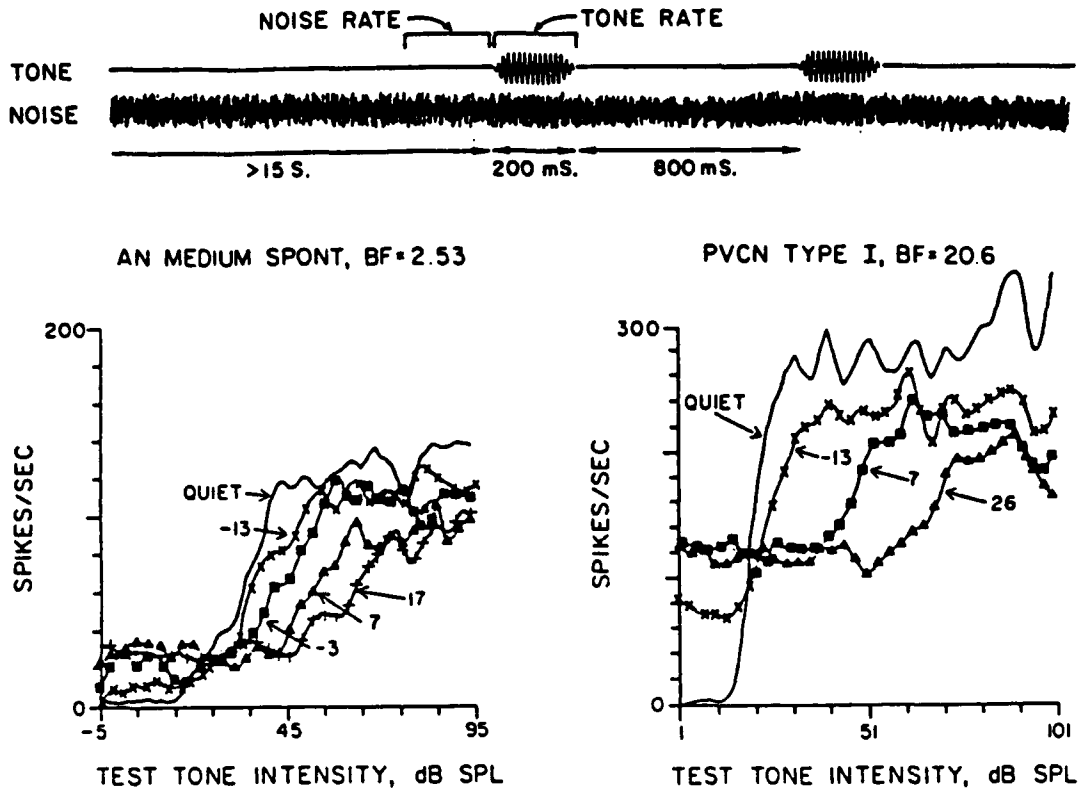


Figure 4 Top line shows stimulus paradigm: a steady broadband noise is turned on for at least 15 s., after which 200 ms. best-frequency tone bursts are presented every second. Left plot shows rate versus tone level for an AN fiber in quiet and four levels of background noise (noise spectrum levels given as parameters on curves). Right plot shows similar data for a unit in the posterior VCN. This unit was not classified in terms of PST histograms. Type I means the unit has no inhibitory sidebands. Reproduced from ref 19.

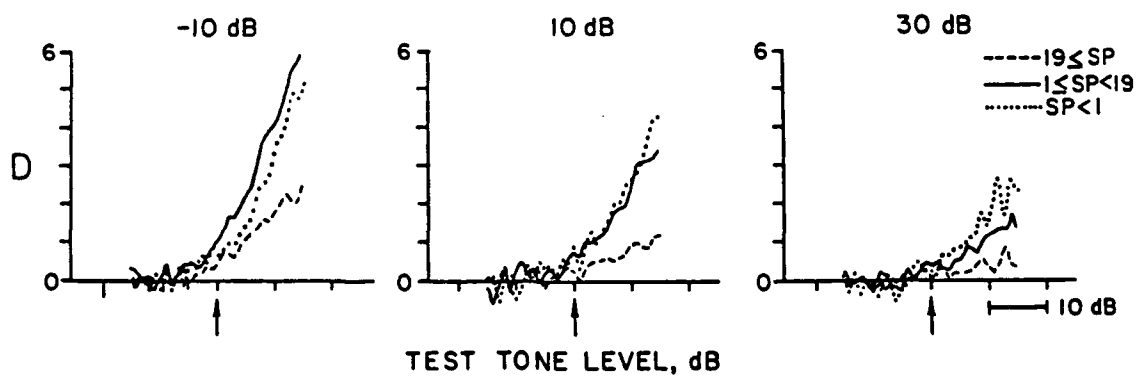


Figure 5 Rate change produced by a tone in the presence of a steady noise background in AN fibers. Abscissae show tone level plotted relative to behavioral masked threshold (arrows). Ordinates show rate change (rate to tone+noise minus rate to noise alone) plotted as D = (rate change)/(rate standard deviation) corrected for effects of the change in rate standard deviation as rate increases (see ref 63 for details). Data averaged across populations of 7-22 fibers. Each plot shows data at one noise level (-10, 10, or 30 db spectrum level). Data for different spontaneous rate groups shown separately, see legend. Reproduced from ref 63.

although there are small quantitative differences in the degree of horizontal shift between the AN and CN (19).

Assume that detection of the tone in the presence of noise is based on detecting the rate increase caused by the tone over the rate response to the noise alone. A convenient way to analyze the tone-driven rate increase in relationship to behavioral masked threshold is to plot rate change in terms of detectability, measured in a way analogous to d' defined above. Fig. 5 shows results for AN fibers (63). These plots show rate change plotted against tone level at three levels of background noise; data are averaged across populations of from 7 to 22 fibers. Rate change is plotted in terms of a variable D which is rate change in units of the standard deviation of rate. D is slightly different from d' defined above because σ_R changes with rate in AN fibers. However, the differences between D and d' should be small in practice. Tone level is plotted on the abscissae in Fig. 5 as tone level relative to the behavioral masked threshold of cats tested under similar stimulus conditions (15); the arrows show the behavioral masked thresholds. Note that the detectability of rate changes begins to increase in all three spontaneous rate groups (defined in the legend) within a few dB of the behavioral masked threshold, suggesting that this analysis is reasonable.

As the noise level increases, there is a noticeable flattening (decrease in slope) of the detectability curves. This flattening is caused by saturation of the fibers' responses due to their response to the noise; the saturation is partially compensated by the horizontal shift in dynamic range shown in Fig. 4. At all noise levels, low and medium spontaneous rate fibers give the most detectable rate changes. The mechanisms responsible for these changes are discussed in the original papers (16,63). What is of interest in this paper is a comparison of these results with results obtained in the CN.

Fig. 6 shows a similar analysis averaged across three regularly-discharging VCN units. Only one of the units, a chopper, was typed using PST histograms, but from their response properties and regularity, it is likely that the other two units were also choppers (51). The results in Fig. 6 are similar to those in Fig. 5 except that D increases more rapidly with tone level in the VCN units of Fig. 6, especially at high noise levels. For comparison, the asterisk in Fig. 6 shows the maximum value of D attained by AN fibers at 30 dB noise level (by the low spontaneous rate fibers). Comparison of the asterisk and the VCN data for 29-39 dB noise shows that the VCN units attain a level of detectability for tone levels 15 dB above masked threshold that is more than twice as large as that of AN fibers. There does not appear to be a substantial shift of the tone level at which rate begins to increase in the VCN units, although there is really no quantitative way to judge this question. However, the VCN units reach any given criterion level of D at a lower tone level than AN fibers.

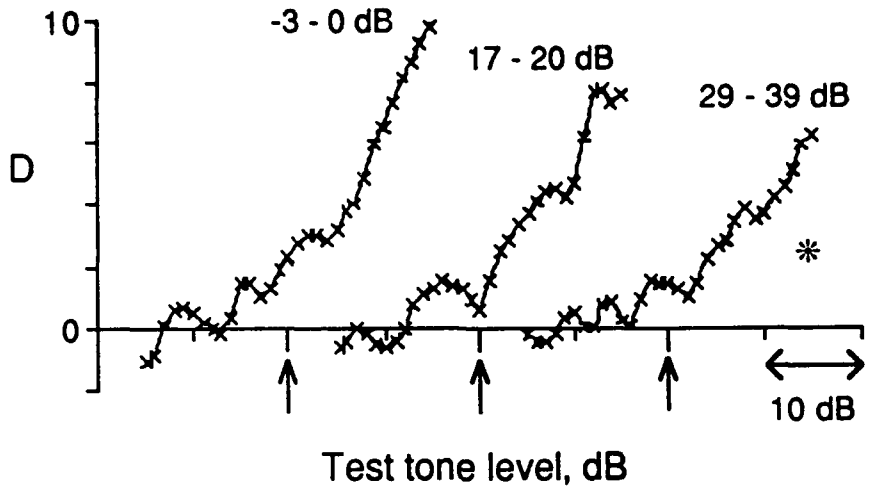


Figure 6 Rate change produced by a tone in the presence of a steady noise background, averaged over three VCN units. Same format as Fig. 5: abscissa shows tone level re behavioral masked threshold (15; arrows) and ordinate shows rate change (rate to tone+noise minus rate to noise alone) plotted in units of one standard deviation of rate. Results averaged across slightly different noise background levels in different units; range of noise levels given next to the curves. Asterisk shows maximum D value achieved by low spontaneous rate AN fibers at 30 dB noise level (from Fig. 5). Method used to compensate for change in standard deviation of rate as rate increases is similar to that used in Fig. 5 (see ref. 63). Statistical model relating mean (μ_N) and standard deviation (σ_N) of spike counts during an interval T is similar to constant-deadtime model of ref 63 and is given by:

$$\sqrt{\frac{\sigma_N^2}{\mu_N}} = a\left[1 - \frac{t_D \mu_N}{T}\right]$$

where t_D is deadtime and a is an additional parameter needed for VCN units. Using this model, detectability D is related to spike count N_T as

$$D = \frac{1}{a\sqrt{t_D/T}} \ln \frac{1 + \sqrt{t_D N_T/T}}{1 - \sqrt{t_D N_T/T}}$$

D has variance approximately 1. The derivation of these two equations follows derivations given in ref 63.

The results of Figs. 5 and 6 show that, all other factors being equal, the detectability of the rate change produced by a tone in noise is higher in regular (chopper) units than in AN fibers (and, by extension, primarylike units). In this sense, the detectability of the tone is enhanced in the chopper unit. Because the slope of the D vs tone level function is higher, it also follows that choppers should support smaller jnds for intensity at suprathreshold levels. Shofner (52) has reported a similar conclusion based on ROC analysis of spike trains from various classes of CN neurons.

Timing and Rate/Intensity Channels in the VCN?

The discussion of the previous section leads naturally to one of the most difficult problems that confronts the analysis of neural responses in the CNS. That problem is how to determine the role that a particular cell type plays in audition, i.e. what sort of information processing is the cell doing? A number of changes in the representation of sound may occur between the inputs and the output of a cell, but it is not clear, from a delineation of those changes, which aspects of the stimulus transformations are actually important in auditory information processing and which aspects are merely epiphenomena. Inferences about a cell's functional role in the auditory nervous system can be made on the basis of its response properties, but solid evidence is hard to come by. It does not follow from the analysis of the previous section, for example, that CN choppers are a channel specialized for representation of stimulus intensity. However, good evidence for this idea has been obtained for the sound localization system of the barn owl (33).

Before considering the barn owl, however, it is necessary to review one additional property of CN units. Fig. 7 shows a comparison of the phase locking abilities of the principal response types in the CN (7). Phase-locking is strong for stimulus frequencies up to 1 kHz in prepotential units (a subset of primarylike units which have prepotentials in their action potentials, 7,40) and in onset units and then gradually drops off to zero by 5 kHz. Blackburn and Sachs (3) show the same behavior for all primarylike units, with or without prepotentials. Phase-locking in primarylike and onset units is essentially the same as is observed in AN fibers (not shown in Fig. 7; see ref 27). By contrast, phase locking in choppers begins to decrease in strength at about 200 Hz and is gone by 2-4 kHz. It is clear from these data that primarylike units provide better information than choppers about stimulus waveform and stimulus phase over most of the frequency range.

From the data in Figs. 6 and 7 and the arguments above, it might be concluded that chopper units constitute a channel for the frequency-specific representation of stimulus power, whereas primarylike units constitute a complementary channel for temporal information about stimulus phase. There is evidence for this idea in the barn owl (56), whose CN contains neurons very similar to mammalian primarylike and chopper units. The barn owl uses interaural phase difference as a cue for azimuthal sound localization and interaural intensity difference as a cue for vertical sound localization (33). The primarylike units of the barn owl are located in a physically different region of the brainstem than the chopper units. Takahashi et al (57) injected lidocaine into these two neural systems separately in order to inactivate them. Injections were made while recording from spatially selective neurons in the midbrain. When the primarylike region was anesthetized, the azimuthal (phase difference dependent) sensitivity of the midbrain neurons was disrupted without affecting the elevational (intensity difference dependent) sensitivity of the cells. When the chopper region was

anesthetized, the opposite result was obtained. These results support the notion of a separate representation of stimulus spectrum and stimulus waveform, or phase, in chopper and primarylike neurons.

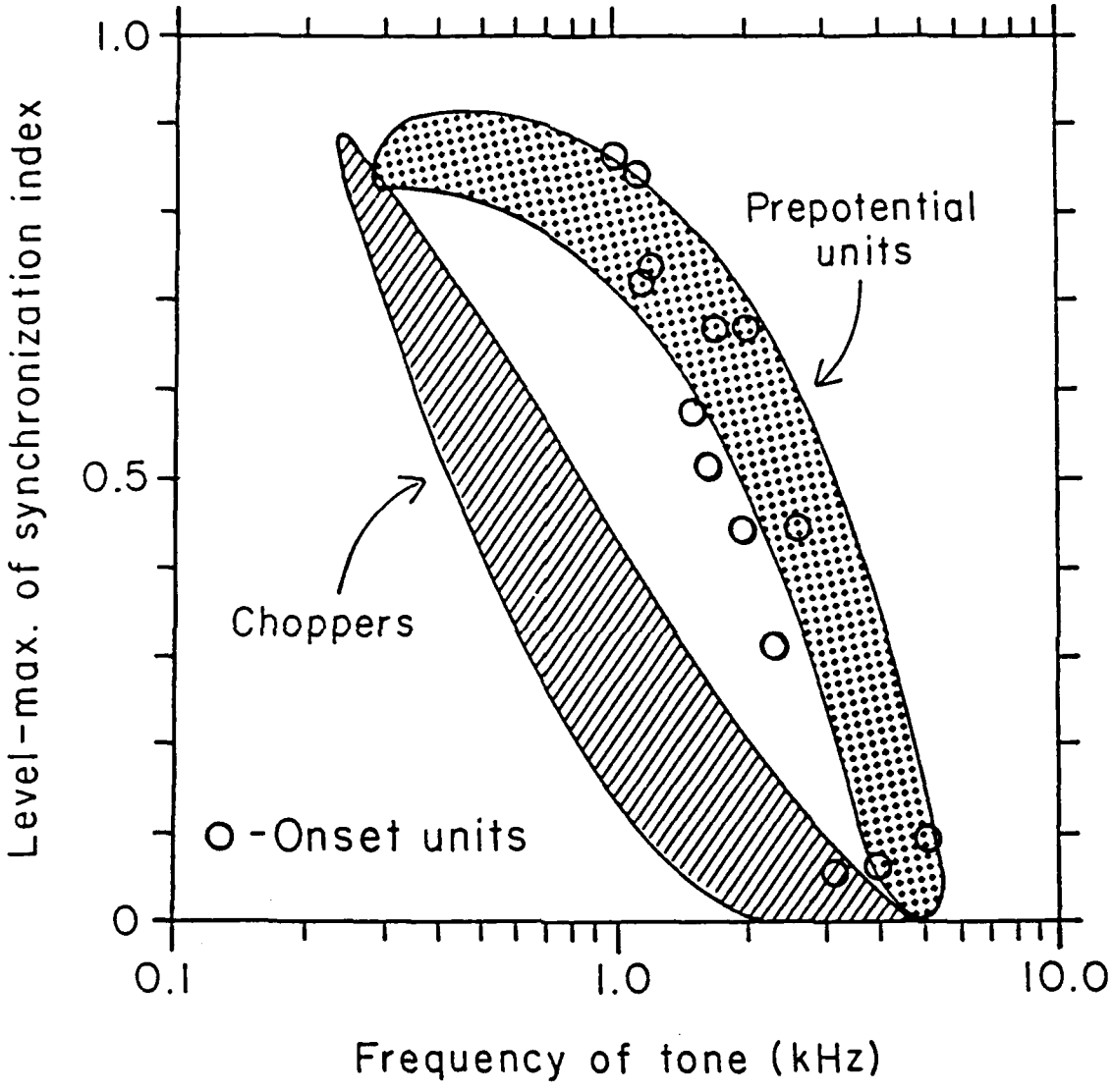


Figure 7 Strength of phase locking versus tone frequency for three major unit types in VCN. Ordinate is maximum synchronization index (or vector strength) observed at any sound level in a unit at a particular frequency. 1 is perfect phase locking (all spikes at the same phase) and 0 is random discharge with respect to stimulus phase. Prepotential units are a subset of primarylike units. Hatched and dotted regions include all data points for chopper and prepotential units; data points for onset units shown. Redrawn from ref 7.

The preceding discussion and the discussion of the next section treat masked threshold, intensity discrimination, and binaural intensity difference sensitivity as derived from the same peripheral representation. This is done to simplify the arguments and seems like a reasonable

assumption at this point, because there is no obvious difference among these three functions in terms of the requirements they place on their peripheral, monaural representations. A similar assumption is made about temporal coding for binaural phase sensitivity and the temporal representation of stimulus spectrum, discussed below.

While the idea of a spectrum and a phase channel at the output of the CN is consistent with a number of results obtained in mammalian systems, it is clear that this idea is an oversimplification in mammals. The anatomical separation of function in the primarylike and chopper systems cannot be as complete in the mammal as the results of Takahashi et al. imply for the owl. Referring to Fig. 1, note that the bushy cells (primarylike responses) of the cat CN are the primary known source of inputs to the principal nuclei of the superior olive, the MSO and LSO (11,60,61; recall that the MNTB is an inhibitory relay to the contralateral LSO). Stellate cell axons (chopper responses) do not seem to terminate in the principal olivary nuclei, although this question has not been settled. The MSO and LSO are believed to be the sites of the initial interaural stimulus comparisons necessary for sound localization. The MSO is a predominantly low frequency nucleus (24) containing cells that are sensitive to interaural phase (21,26). The preservation of stimulus phase information in CN primarylike axons is clearly necessary for the interaural phase sensitivity of the MSO. The LSO, by contrast, is a predominantly high frequency nucleus (6,24) whose cells are sensitive to interaural intensity differences (6,26) and perhaps also interaural time delays of the stimulus envelope (26). Because bushy cells are a major source of input to the LSO in the cat, the information for binaural intensity comparisons is provided by primarylike units of the CN, so that primarylike units convey intensity as well as timing information in the mammal.

The example of the mammalian LSO illustrates another important aspect of the analysis of information processing in the CNS: it is not possible to critically test hypotheses about the functional role of a neuron type without a knowledge of its projective field, i.e. the destinations of its axons. Another illustration of this fact is provided by the dorsal cochlear nucleus, where a substantial fraction of the neurons are interneurons. It is clearly essential, in trying to interpret the response properties of a DCN unit, to know whether it is a principal cell that constitutes the output of the nucleus or an interneuron which participates in the internal processing within the nucleus.

Complex Stimuli

The stimuli that the auditory system normally encounters have complex, time-varying spectra. These natural stimuli are very different from the kinds of simple laboratory stimuli discussed above. So far, the greatest progress in understanding the central auditory system has come from analyses of natural stimulus situations, where the system's tasks are clearly known. The best example of this is

the analysis of the processing of bat sonar signals in the cortex (see ref 55 for a review). Sound localization systems are another example (26,33,61A). Although simple laboratory signals turn out to be useful in analyzing sound localization, our ultimate understanding of information processing in the auditory system will require the use of complex, natural stimuli. Complex stimuli are necessary for two major reasons. First, questions about the functional relevance of a particular stimulus transformation can be most clearly stated and answered in natural stimulus situations where the behavioral relevance of the stimuli can be specified. Using natural stimuli is the only general way around the problem, posed above, of what the functional role of a particular cell is. Second, there are likely to be special, highly non-linear, mechanisms in the system for the analysis of certain special stimulus situations. These mechanisms may be impossible to study with simple laboratory stimuli, either because the responses are so complex that they do not suggest clear hypotheses, or because the units do not respond to simple stimuli. Examples of such special mechanisms are provided by the combination-sensitive units of the bat cortex (55) and the monkey-face and monkey-hand selective units in the insulo-temporal cortex (17).

The best studied complex auditory signal is human speech (see ref 49 for review). Although speech is not a natural stimulus for the cat, the study of responses to speech in the CN can be used to illustrate some of the points raised above. Fig. 8 shows data of Blackburn and Sachs on responses to a steady-state synthetic /ɛ/ in a primarylike unit (left column) and a chopper (right column). The units' best frequencies are near the second formant frequency of the vowel (1792 Hz). Note, in the period histograms in the second row, that the primarylike unit is responding in a phase-locked manner to the second formant component of the vowel (16^{th} harmonic), whereas the chopper is not. The chopper's response is primarily phase locked to the stimulus fundamental, with only a weak response to stimulus energy near its BF. The nature of these responses can be clearly seen in the Fourier transform plots in the bottom row.

The results in Fig. 8 are expected, on the basis of the phase-locking data in Fig. 7. These data are typical of the responses of populations of primarylike and chopper units (4) and are consistent with the idea that primarylike units constitute a channel conveying temporally-coded information about stimulus phase and stimulus waveform, whereas chopper units do not carry stimulus phase information.

Chopper units do convey a good representation of the power spectrum of the stimulus, as the data in Fig. 9 show. Fig. 9 compares profiles of discharge rate versus place (i.e. best frequency) for high spontaneous rate AN fibers (top plot), chop-S units (middle plot) and chop-T units (bottom plot). In each case, the stimulus is the same steady-state synthetic /ɛ/. The chopper populations provide a robust representation of the stimulus spectrum, in that peaks of response are associated with

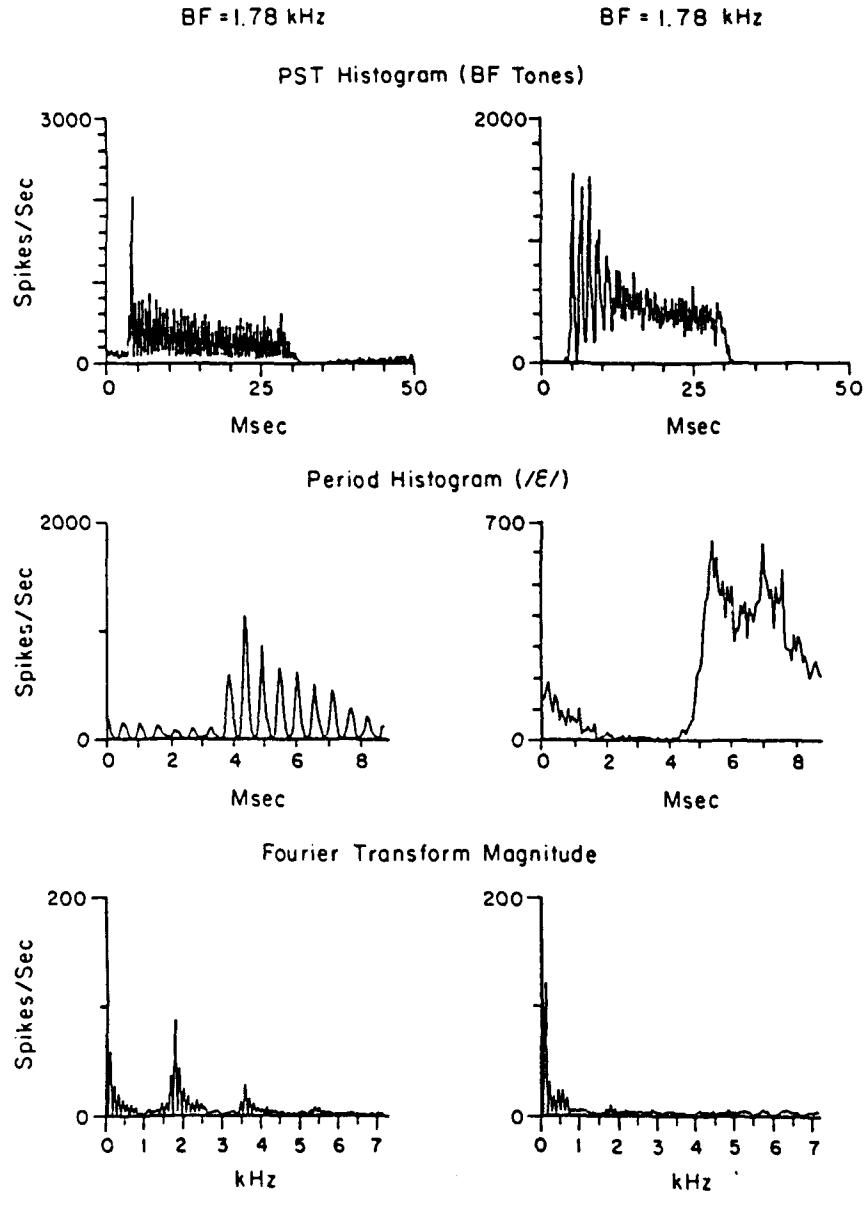


Figure 8 Examples of responses to a steady-state vowel /ɛ/ in two CN neurons. The left column shows data from a primarylike neuron and the right column shows data from a chopper. Top row: PST histograms of responses to best-frequency tones. Second row: period histograms of responses to the vowel. Bottom row: magnitude spectrum of Fourier transforms of the period histograms. Units' best frequencies are near the second formant frequency of the vowel, equal to its 16th harmonic. The fundamental frequency of the periodic vowel is 112 Hz. Reproduced from ref 49.

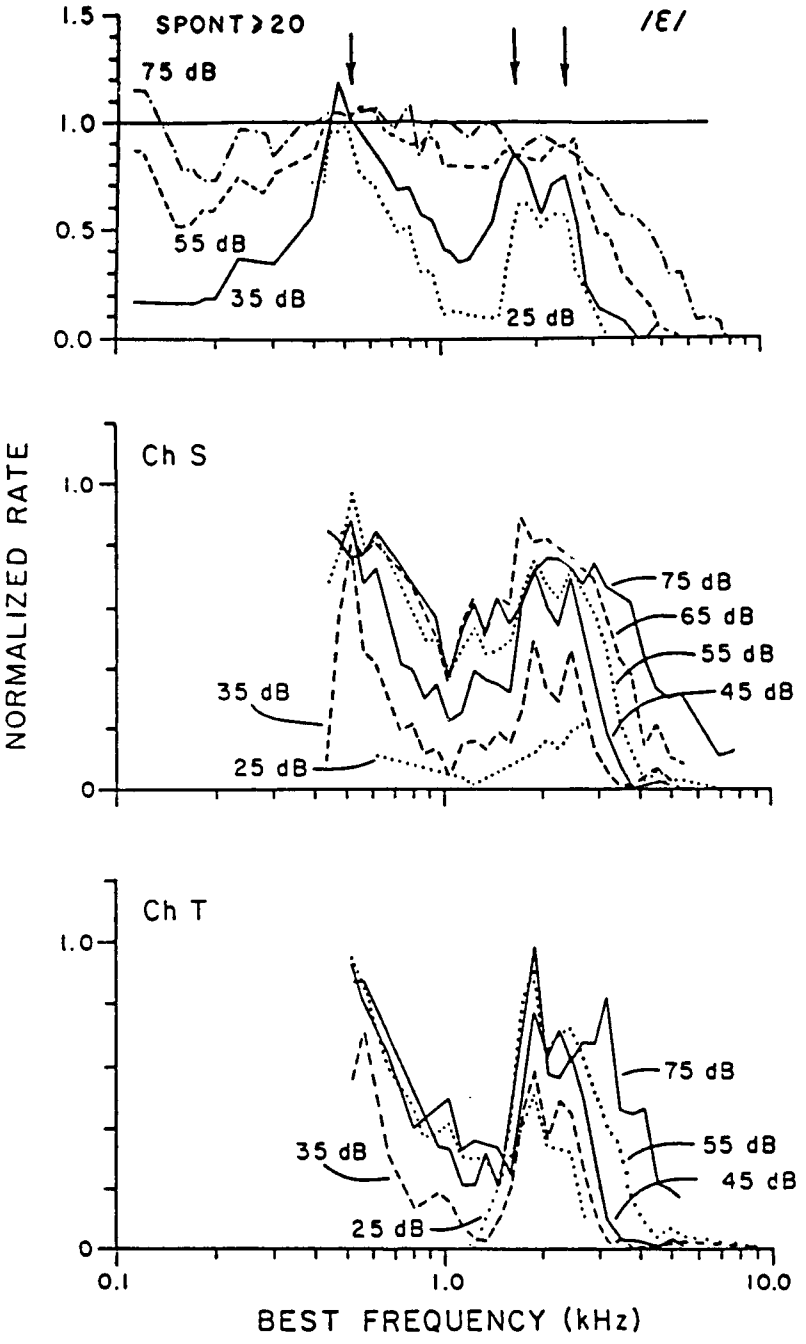


Figure 9 Plots of discharge rate versus place (best frequency) for populations of high spontaneous rate AN fibers (top plot), CN chop-S units (middle plot) and CN chop-T units (bottom plot). Rate is normalized on the abscissa so that 0 is spontaneous rate and 1 is saturation rate to a best-frequency tone. The lines in these plots are moving-window averages of actual data points (see ref 50 for details). Few low best frequency chopper units were encountered, so profiles are cut off on low frequency side. Parameters on the plots show SPL of the vowel. Arrows show formant frequencies. AN data from ref 50; CN data from ref 4.

the formants at all stimulus levels. By contrast, the response of the high spontaneous rate AN population is not robust, in that the representation degenerates at high stimulus levels as fibers between the first two formants saturate (recall that low and medium spontaneous rate fibers provide a somewhat better representation at high levels; 50).

Primarylike units also convey a good representation of the stimulus spectrum (4; data not shown). The clearest representation is carried in a temporal form revealed by phase-locking analyses such as the ALSR (49). Discharge rate profiles for primarylike units behave similarly to those of AN fibers, but with much more variability in response. Thus, the idea of a spectrum channel (choppers) and a phase channel (primarylike units) at the output of the CN has to be qualified still further to incorporate the fact that primarylike units provide excellent information about stimulus spectrum in a temporally-encoded form.

The data in Fig. 10 provide an example of an unexpected result observed with the complex vowel stimulus. This figure shows data of Blackburn and Sachs from two primarylike units and one onset unit. The left-hand dot display for each unit shows responses to a series of best-frequency tones over a range of sound levels. The response types of the three units are clear from these dot displays. The right-hand dot displays show responses to the vowel /ɛ/ and the plots at right show rate-versus-level functions for responses to best-frequency (BF) tones and the vowel. For the primarylike unit (top row), the response to the vowel is somewhat weaker than the response to a BF tone. This is the behavior observed in AN fibers and is expected from suppression by the first formant energy in the vowel (50).

By contrast, the responses of the other two units to the vowel are much stronger than their responses to BF tones. Inspection of the dot displays shows that the units are responding in a one-to-one fashion to the stimulus fundamental. Notice the precise vertical rows of dots and the mean discharge rates very near the fundamental frequency (112 Hz) over a range of 30-40 dB, beginning about 10 dB above threshold. These responses are typical of onset units and some pri-N units and are not expected from AN fiber data, given the best frequencies and phase locking characteristics of the units. The responses shown in Fig. 10 represent some specialization, not yet fully understood, which may signal sudden changes in stimulus amplitude like those which occur each cycle of the vowel. Results of this kind have been previously reported by Kim et al. (32).

Pinna-Filtered Stimuli

Fig. 11 shows three examples of transfer functions from a speaker in the free field (see legend for positions) to a point near the tympanic membrane of an anesthetized cat (44; similar results are

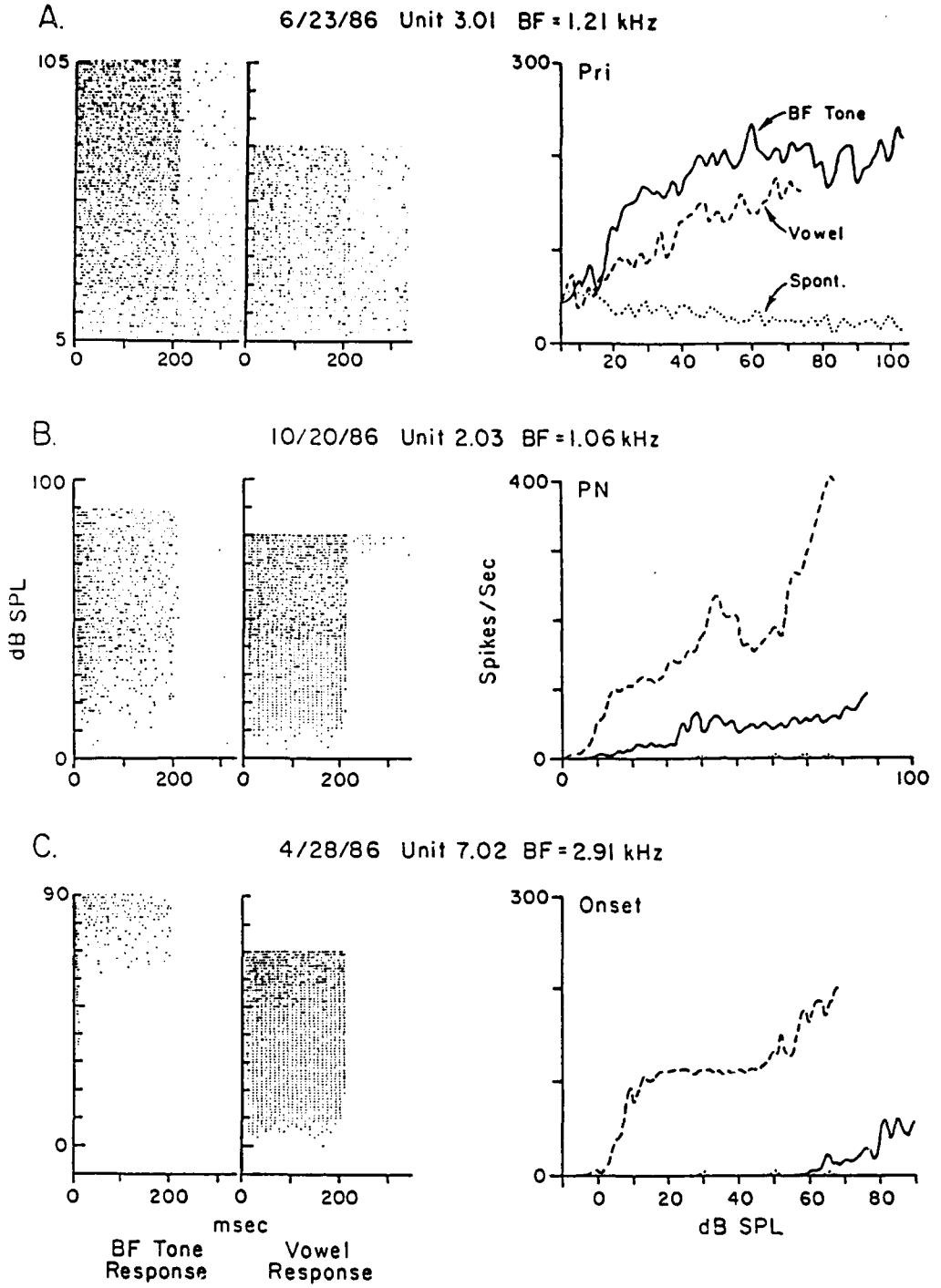


Figure 10 Responses to best-frequency tones and to /ɛ/ of a primarylike unit (top), a primarylike with notch unit (middle), and an onset unit (bottom). Dot displays at left show responses to best-frequency tones (left dot display) and to the vowel (right dot display) over a range of sound levels, as labelled on the ordinates. Each dot display shows responses to 100 successive presentations of tone or vowel, with the stimulus level increasing by 1 dB between stimulus presentations. Plots at right show rate versus sound level for best-frequency tone (solid lines), vowel (dashed line), and spontaneous activity (dotted line). Reproduced from ref 4.

reported in ref 36). The prominent spectral features at frequencies above 5 kHz are due to directionally dependent filtering by the external ear and provide strong cues for sound localization. One cue, for example, is the frequency of the spectral notch near 10 kHz, called the first notch (FN). The FN frequency increases monotonically with azimuth and elevation and provides a direct cue for spatial location in the frontal field. A requirement for use of this cue is that the stimulus be broad band or that the animal be familiar with the stimulus, so that spectral features of the stimulus can be separated from those contributed by the external ear. These requirements are consistent with psychophysical data on the use of similar cues by human observers (see ref 5 for a review).

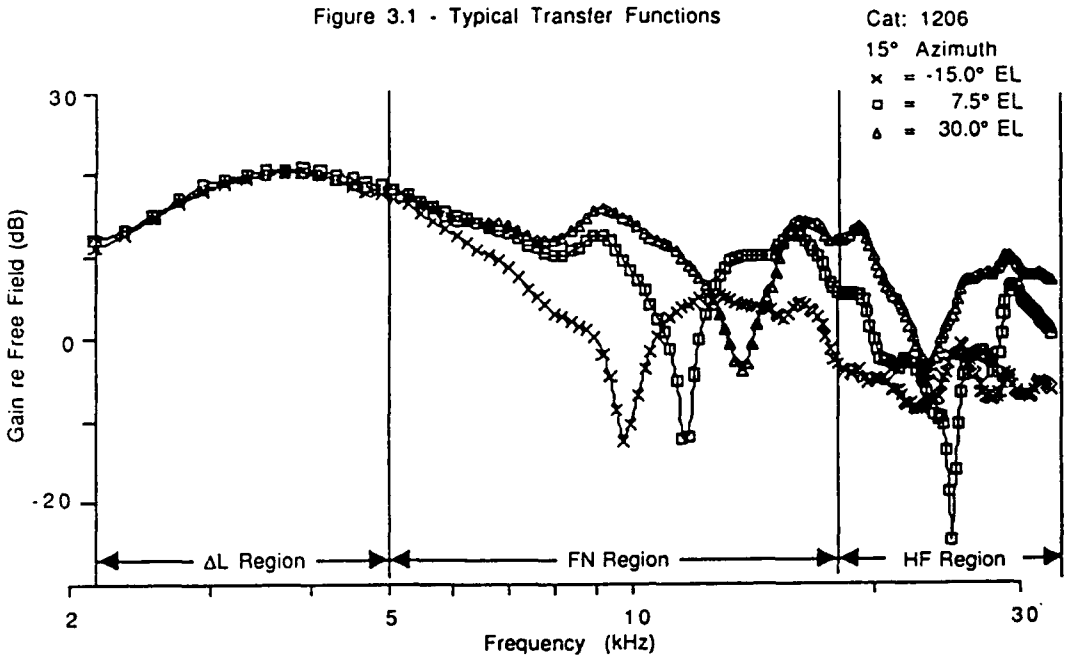


Figure 11 Transfer functions from a speaker at three positions in the free field to a point near the tympanic membrane. Positions are given in the legend: 0° azimuth and 0° elevation is straight in front of the cat. Positive azimuths are toward the ear in which the transfer function is being measured and positive elevations are upward. Gain is given re the transfer function to a probe tube in the free field with the cat removed. Data from ref 44.

The use of cues in pinna-filtered stimuli is an example of a perceptual problem that is amenable to physiological and psychophysical study in the same species. Pinna-filtered stimuli have the advantage that they are a natural stimulus, which should aid the design and interpretation of physiological experiments. Stimuli such as pinna filtered sounds may provide a basis for exciting new ventures in physiological/psychophysical correlation.

Acknowledgements

Unpublished data on CN responses in Figs. 2, 9 and 10 were provided by Carol Blackburn and Murray Sachs. Comments of Carol Blackburn, Murray Sachs, and George Spirou on this manuscript are appreciated.

Preparation of this paper was supported by grant NS12524 from the NIH and by a conference grant from the AFOSR.

References

1. Adams, J.C. Ascending projections to the inferior colliculus. *J. Comp. Neurol.* 183: 519-538, 1979.

2. Adams, J.C. Multipolar cells in the ventral cochlear nucleus project to the dorsal cochlear nucleus and the inferior colliculus. *Neurosci. Lett.* 37:205-208, 1983.

3. Blackburn, C.C. and Sachs, M.B. Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *J. Neurophysiol.*, in press, 1989.

4. Blackburn, C.C. and Sachs, M.B. The representation of the steady-state vowel /ɛ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. In preparation.

5. Blauert, J. *Spatial Hearing* The MIT Press, Cambridge, Massachusetts, 1983.

6. Boudreau, J.C. and Tsuchitani, C. Cat superior olive S-segment cell discharge to tonal stimuli. In: *Contributions to Sensory Physiology*, edited by W.D. Neff, New York: Academic Press. 4:143-213, 1970.

7. Bourk, T.R. *Electrical Responses of Neural Units in the Anteroventral Cochlear Nucleus of the Cat.* (Doctoral dissertation). Cambridge: Massachusetts Institute of Technology, 1976.

8. Brawer, J.R., Morest, D.K. and Kane, E.C. The neuronal architecture of the cochlear nucleus of the cat. *J. Comp. Neurol.* 155: 251-300, 1974.

9. Cant, N.B. The fine structure of two types of stellate cells in the anterior division of the anteroventral cochlear nucleus of the cat. *Neuroscience* 6:2643-2655, 1981.

10. Cant, N.B. Identification of cell types in the anteroventral cochlear nucleus that project to the inferior colliculus. *Neurosci. Lett.* 32: 241-246, 1982.

11. Cant, N.B. and Cassedy, J.H. Projections from the anteroventral cochlear nucleus to the lateral and medical superior olivary nuclei. *J. Comp. Neuro.* 247:457-476, 1986.

12. Cant, N.B. and Gaston, K.C. Pathways connecting the right and left cochlear nuclei. *J. Comp. Neurol.* 212:313-326, 1982.

13. Cant, N.B. and Morest, D.K. The bushy cells in the anteroventral cochlear nucleus of the cat. A study with the electron microscope. *Neurosci.* 4:1925-1945, 1979.

14. Cant, N.B. and Morest, D.K. The structural basis for stimulus coding in the cochlear nucleus of the cat. In: *Hearing Science Recent Advances*, edited by C.I. Berlin. San Diego: College Hill Press, 1984. p. 371-422.

15. Costalupes, J.A. Broadband masking noise and behavioral pure tone thresholds in cats. *J. Acoust. Soc. Am.* 74:758-764, 1983.

16. Costalupes, J.A., Young, E.D., and Gibson, D.J. Effects on continuous noise backgrounds on rate response of auditory nerve fibers in cat. *J. Neurophysiology* 51:1326-1344, 1984.

17. Desimone, R., Albright, T.D., Gross, C.G. and Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4:2051-2062, 1984.

18. Evans, E.F. and Wilson, J.P. The frequency selectivity of the cochlea. In: *Basic Mechanisms in Hearing*, edited by A.R. Møller, New York: Academic Press, p. 519-554, 1973.

19. Gibson, D.J., Young, E.D., and Costalupes, J.A. Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei. *J. Neurophysiol.* 53:940-958, 1985.

20. Godfrey, D.A., Kiang, N.Y.S., and Norris, B.E. Single unit activity in the posteroventral cochlear nucleus of the cat. *J. Comp. Neurol.* 162: 247-268, 1975.

21. Goldberg, J.M. and Brown, P.B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. *J. Neurophysiol.* 32:613-636, 1969.

22. Goldberg, J.M. and Greenwood, D.D. Response of neurons of the dorsal and posteroventral cochlear nuclei of the cat to acoustic stimuli of long duration. *J. Neurophysiol.* 29:72-93, 1966.

23. Green, D.M. and Swets, J.A. *Signal Detection Theory and Psychophysics.* New York: Krieger, 1974.

24. Guinan, J.J. Jr., Norris, B.E. and Guinan, S.S. Single auditory units in the superior olivary complex II: Locations of unit categories and tonotopic organization. *Intern. J. Neuroscience* 4:147-166, 1972.

25. Hirsh, J.A. and Oertel, D. Intrinsic properties of neurones in the dorsal cochlear nucleus of mice, *in vitro. J. Physiol.(Lond.)* 396:535-548, 1988.

26. Irvine, D.R.F. The auditory brainstem. *Progress in Sensory Physiology*, volume 7, edited by D. Ottoson, Berlin: Springer-Verlag, 1986.

27. Johnson, D.H. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.* 68: 1115-1122, 1980.

28. Kane, E.C. Octopus cells in the cochlear nucleus of the cat: heterotypic synapses upon homeotypic neurons. *Intern. J. Neurosci.* 5:251-279, 1973.

29. Kane, E.C. Synaptic organization in the dorsal cochlear nucleus of the cat: A light and electron microscopic study. *J. Comp. Neurol.* 155:301-330, 1974.

30. Kiang, N.Y.S., Morest, D.K., Godfrey, D.A., Guinan, J.J., and Kane, E.S. Stimulus coding at caudal levels of the cat's auditory system: I. Response characteristics of single units. In: *Basic Mechanisms in Hearing*, edited by A.R. Møller. New York: Academic Press, p.455-478, 1973.

31. Kiang, N.Y.S., Watenabe, T., Thomas, E.C. and Clark, L.F. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve.* Cambridge, MA.: MIT Press, 1965.

32. Kim, D.O., Rhode, W.S., and Greenberg, S.R. Responses of cochlear nucleus neurons to speech signals: neural encoding of pitch, intensity and other parameters. In: *Auditory Frequency Selectivity*, edited by B.C.J. Moore and R.B. Patterson. New York: Plenum, p 281-288, 1986.

33. Konishi, M., Takahashi, T.T., Wagner, H., Sullivan, W.E., and Carr, C.E. Neurophysiological and anatomical substrates of sound localization in the owl. In: *Auditory Function: Neurobiological Bases of Hearing*, edited by G.M. Edelman, W.E. Gall, and W.M. Cowan. New York: John Wiley. p. 721-745, 1988.

34. Liberman, M.C. Auditory-nerve response from cats raised in a low-noise chamber. *J. Acoust. Soc. Am.* 63: 442-455, 1978.

35. Lorente de No, R. *The Primary Acoustic Nuclei.* New York:Raven Press, 1981.

36. Musicant, A.D., Chan, J.C.K. and Hind, J.E. Distribution of sound pressure along the surface of cat tympanic membrane. *J. Acoust. Soc. Am.* 77:S94 (abstract), 1985.

37. Oertel, D. Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus. *J. Neuroscience* 3:2043-2053, 1983.

38. Osen, K.K. Cytoarchitecture of the cochlear nuclei in the cat. *J. Comp. Neurol.* 136: 453-484, 1969.

39. Osen, K.K. and Mugnaini, E. Neuronal circuits in the dorsal cochlear nucleus. In: *Neuronal Mechanisms of Hearing,* edited by J. Syka and L. Aitkin. New York: Plenum. p.119-125, 1981.

40. Pfeiffer, R.R. Anteroventral cochlear nucleus: Waveforms of extracellularly recorded spike potentials. *Science* 134: 667-668, 1966a.

41. Pfeiffer, R.R. Classification of response patterns of spike discharges for units in the cochlear nucleus: Tone burst stimulation. *Exp. Brain Res.* 1: 220-235, 1966b.

42. Rhode, W.S. and Smith, P. H. Encoding timing and intensity in the ventral cochlear nucleus of the cat. *J. Neurophysiol.* 56:26l-286, 1986

43. Rhode, W.S., Oertel, D. and Smith, P.H. Physiological response properties of cells labeled intracellularly with horseradish peroxidase in the cat ventral cochlear nucleus. *J. Comp. Neurol.* 213:448-463, 1983.

44. Rice, J.J., May, B.. Spirou, G.A., and Young, E.D. Directional dependence of free-field-to-eardrum transformations in cats. *J. Acoust. Soc. Am.* 85:S67, 1989 (abstract).

45. Ritz, L.A. and Brownell, W.E. Single unit analysis of the posteroventral cochlear nucleus of the decerebrate cat. *Neuroscience* 7: 1995-2010, 1982.

46. Romand, R. Survey of intracellular recording in the cochlear nucleus of the cat. *Brain Res.* 148:43-65, 1978.

47. Rouiller, E.M. and Ryugo, D.K. Intracellular marking of physiologically characterized cells in the ventral cochlear nucleus of the cat. *J. Comp. Neurol.* 225:167-186, 1984.

48. Ryugo, D.K. and Fekete, D.M. Morphology of primary axosomatic endings in the anteroventral cochlear nucleus of the cat: A study of the endbulbs of held. *J. Comp. Neurol.* 120:239-257, 1982.

49. Sachs, M.B., Winslow, R.L. and Blackburn, C.C. Representation of speech in the auditory periphery. In: *Auditory Function: Neurobiological Bases of Hearing*, edited by G.M. Edelman, W.E. Gall, and W.M. Cowan. New York: John Wiley. p. 747-774, 1988.

50. Sachs, M.B. and Young, E.D. Encoding of steady-state vowels in the auditory nerve: Encoding in terms of discharge rate. *J. Acoust. Soc. Am.* 66:470-479, 1979.

51. Shofner, W.P. and Young, E.D. Excitatory/inhibitory response types in the cochlear nucleus: Relationships to discharge patterns and responses to electrical stimulation of the auditory nerve. *J. Neurophysiol.* 54:917-939, 1985.

52. Shofner, W.P. Intensity discrimination by cochlear nucleus neurons: ROC analysis of empirical spike count distributions. *Abst. Midwinter Mtg. of the Assoc. for Res. in Otolaryngol.* 12:60, 1989 (abstract).

53. Siebert, W.M. Stimulus transformations in the peripheral auditory system. In: *Recognizing Patterns*, edited by P.A. Kolers and M. Eden. Cambridge: MIT Press, p. 104-133, 1968.

54. Smith, P.H. and Rhode, W.S. Characterization of HRP-labeled globular bushy cells in the cat anteroventral cochlear nucleus. *J. Comp. Neurol.* 266:360-376, 1987.

55. Suga, N. Auditory neuroethology and speech processing: Complex-sound processing by combination-sensitive neurons. In: *Auditory Function: Neurobiological Bases of Hearing*, edited by G.M. Edelman, W.E. Gall, and W.M. Cowan. New York: John Wiley. p. 679-720, 1988.

56. Sullivan, W.E. and Konishi, M. Neural map of interaural phase difference in the owl's brainstem. *Proc. Natl. Acad. Sci.* 83:8400-8404, 1986.

57. Takahashi, T.T., Moiseff, A., and Konishi, M. Time and intensity cues are processed independently in the auditory system of the owl. *J. Neurosci.* 4:1781-1786, 1984.

58. Tolbert, L.P. and Morest, D.K. The neuronal architecture of the anteroventral cochlear nucleus of the cat in the region of the cochlear nerve root: Golgi and Nissl methods. *Neuroscience* 7: 3013-3030, 1982.

59. Tolbert, L.P. and Morest, D.K. The neuronal architecture of the anteroventral cochlear nucleus of the cat in the region of the cochlear nerve root: Electron microscopy. *Neuroscience* 7:3053-3067, 1982.

60. Tolbert, L.P., Morest, D.K., and Yurgelun-Todd, D. The neuronal architecture of the anteroventral cochlear nucleus of the cat in the region of the cochlear nerve root: horseradish peroxidase labelling of identified cell types. *Neuroscience* 7: 3031-3052, 1982.

61. Warr, W.B. Parallel ascending pathways from the cochlear nucleus: neuroanatomical evidence of functional specialization. *Contrib. Sens. Physiol.* 7: 1-38, 1982.

61A. Yin, T.C.T. and Chan, J.C.K. Neural mechanisms underlying interaural time sensitivity to tones and noise. In: *Auditory Function: Neurobiological Bases of Hearing*, G.M. Edelman, W.E. Gall, and W.M. Cowan (Eds.), Wiley & Sons, Publ., p. 385-430, 1988.

62. Young, E.D. Response characteristics of neurons of the cochlear nuclei. In: *Hearing Science Recent Advances*, edited by C.I. Berlin. San Diego: College-Hill Press, p. 423-460, 1984.

63. Young, E.D. and Barta, P.E. Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J. Acoust. Soc. Am.* 79:426-442, 1986.

64. Young, E.D., Robert, J.-M. and Shofner, W.P. Regularity and latency of units in ventral cochlear nucleus: Implications for unit classification and generation of response properties. *J. Neurophysiol.* 60: 1-29, 1988.

65. Young, E.D., Shofner, W.P., White, J.A., Robert, J.-M., and Voigt, H.F. Response properties of cochlear nucleus neurons in relationship to physiological mechanisms. In: *Auditory Function: Neurobiological Bases of Hearing*, G.M. Edelman, W.E. Gall, and W.M. Cowan (Eds.), Wiley & Sons, Publ., p. 277-312, 1988.